# Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions

**Michael B. Morrissey**[1,2] (ID) **and I. B. J. Goudie**[3] (ID)

[1]*School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK*

 [2]*E-mail: michael.morrissey@st-andrews.ac.uk*

[3]*School of Mathematics and Statistics, University of St Andrews, St Andrews, Fife KY16 9SS, UK*

**Log-linear models are widely used for assessing determinants of fitness in empirical studies, for example, in determining how reproductive output depends on trait values or environmental conditions. Similarly, theoretical works of fitness and natural selection employ log-linear models, often with a negative quadratic term, generating Gaussian fitness functions. However, in the specific application of regression-based analysis of natural selection, such models are rarely employed. Rather, OLS regression is the predominant means of assessing the form of natural selection. OLS regressions allow specific evolutionary quantitative parameters, selection gradients, to be estimated, and benefit from the fact that the associated statistical models are easily applied. We examine whether selection gradients can be directly expressed in terms of the coefficients of models using exponential fitness functions with linear or quadratic arguments. Such models can be easily fitted with generalized linear models (GLMs). The expressions we obtain coincide with those for Gaussian functions, but relax the major constraint that the (log) fitness function is concave (downwardly curved). Additionally these results lead to univariate and multivariate analyses of both linear and quadratic selection that potentially incorporate pragmatic and interpretable models of fitness functions, where the parameters can be related analytically to selection gradients, and that can be operationalized using widely available statistical tools.**

**KEY WORDS:** Capture-mark-recapture, fitness, generalised linear model, natural selection, selection gradients, survival analysis.

The characterization of natural selection, especially in the wild, has long been a major research theme in evolutionary ecology and evolutionary quantitative genetics (Endler 1986, Kingsolver et al. 2001, Lande & Arnold 1983, Manly 1985, Weldon 1901). In recent decades, regression-based approaches have been used to obtain direct selection gradients (especially following Lande & Arnold 1983), which represent the direct effects of traits on fitness. These, and related, measures of selection have an explicit justification in quantitative genetic theory (Lande 1979, Lande & Arnold 1983), which provides the basis for comparison among traits, taxa, and so on, and ultimately allows meta-analysis (e.g., Kingsolver et al. 2001). Selection gradients can characterize both directional selection and aspects of nonlinear selection, and so are a very powerful concept in evolutionary quantitative genetics.

The selection gradient may be defined as the vector of partial derivatives of relative fitness with respect to phenotype, averaged over the distribution of phenotype observed in a population. This definition is equivalent to other existing definitions when the phenotype follows a Gaussian distribution (Walsh & Lynch 2018), an assumption we rely on for most of our results. Following Lande & Arnold (1983), given an arbitrary function $W(\mathbf{z})$ for expected fitness of a (multivariate) phenotype $\mathbf{z}$, a general expression for the directional selection gradient vector $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = \bar{W}^{-1} \int \frac{\partial W(\mathbf{z})}{\partial \mathbf{z}'} p(\mathbf{z}) d\mathbf{z} \,, \tag{1}$$

where $p(\mathbf{z})$ is the probability density function of phenotype, with $\mathbf{z}$ being a column vector, and $\bar{W}$ is mean fitness. Mean fitness can itself be obtained by $\int W(\mathbf{z})p(\mathbf{z})d\mathbf{z}$. A quadratic selection

gradient can also be defined as the average curvature (similarly standardised), rather than the average slope, of the relative fitness function,

$$\gamma = \bar{W}^{-1} \int \frac{\partial^2 W(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}'} p(\mathbf{z}) d\mathbf{z}. \tag{2}$$

The directional selection gradient has a direct relationship to evolutionary change, assuming that breeding values (the additive genetic component of individual phenotype, Falconer 1960) are multivariate normally distributed, following the Lande (1979) equation

$$\Delta \bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}, \tag{3}$$

where $\Delta \bar{\mathbf{z}}$ is per-generation evolutionary change, and $\mathbf{G}$ is the additive genetic covariance matrix, i.e., the (co)variances among individuals of breeding values. The quadratic selection gradient matrix has direct relationships to the change in the distribution of breeding values due to selection, but not with such simple relationships between generations as for the directional selection gradient and the change in the mean (Lande & Arnold 1983). Walsh & Lynch (2018) provide an extended treatment of the various relationships between the summaries of the local fitness function provided by $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to changes in the distribution of phenotype and breeding values.

Some progress has been made at developing generalised regression model methods for inference of selection gradients. Janzen & Stern (1998) proposed a method for binomial fitness components (e.g., per-interval survival, mated vs. not mated). The Janzen & Stern (1998) method provides estimates of $\boldsymbol{\beta}$, and requires fitting a logistic model with linear terms only, calculating the average derivatives at each phenotypic value observed in a sample, and then standardizing to the relative fitness scale. Morrissey & Sakrejda (2013) expanded Janzen & Stern's (1998) basic approach to arbitrary fitness functions (i.e., not necessarily linear) and arbitrary response variable distributions, retaining the basic idea of numerically averaging the slope (and curvature) of the fitness function over the distribution of observed phenotype. Shaw & Geyer (2010) developed a framework for characterizing the distributions of fitness (and fitness residuals) that arise in complex life cycles, and also showed how the method could be applied to estimate selection gradients by averaging the slope or curvature of the fitness function over the observed values of phenotype in a sample.

Perhaps the simplest fitness function, $W(\mathbf{z})$, that arises in evolutionary theory is a log-linear model, such that

$$W(\mathbf{z}) = e^{\alpha + \mathbf{b}'\mathbf{z}}, \tag{4}$$

where $\boldsymbol{\alpha}$ is an intercept, and $\mathbf{b}$ is a vector of linear regression coefficients on the log scale. The derivative of fitness with respect

to phenotype is $\partial W/\partial \mathbf{z}' = \mathbf{b}W(\mathbf{z})$, and so from equation 1 the selection gradient is

$$\boldsymbol{\beta} = \frac{\int \mathbf{b}W(\mathbf{z})p(\mathbf{z})d\mathbf{z}}{\int W(\mathbf{z})p(\mathbf{z})d\mathbf{z}} = \mathbf{b}. \tag{5}$$

Thus, under such a model, the directional selection gradient vector $\boldsymbol{\beta}$ is equal to $\mathbf{b}$ (see also Lande 1983, Chevin & Hospital 2008, Chevin et al. 2015). The fitness function given by equation 4 is convex (upwardly curved) on the scale of fitness (as opposed to log fitness), and this may be regarded both as a biological feature of such a model, or as a statistical artifact of fitting a model that is linear on the log scale (Schluter 1988, Chevin & Hospital 2008).

More commonly in evolutionary theory (e.g., Lande 1976, Chevin & Haller 2014), Gaussian functions are used, such that

$$W(\mathbf{z}) \propto e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\theta})'\boldsymbol{\omega}^{-1}(\mathbf{z}-\boldsymbol{\theta})}, \tag{6}$$

where $\theta$ is the vector of optimal phenotypes and $\omega$ describes the curvature; $\omega$ must be positive-definite. Any time that mean trait values are not equal to the optimum ($\theta$), selection in a Gaussian fitness model has a directional component. Specifically,

$$\boldsymbol{\beta} = -\mathbf{S}(\boldsymbol{\mu} - \boldsymbol{\theta}), \tag{7}$$

where $\mu$ is mean phenotype and $\mathbf{S} = (\boldsymbol{\omega} + \boldsymbol{\Sigma})^{-1}$, where $\boldsymbol{\Sigma}$ is the phenotypic variance-covariance matrix before selection; this relation is used extensively in evolutionary theory (e.g., Lande 1976, 1979, Gomulkiesicz & Houle 2009, Chevin & Haller 2014). Quadratic selection gradients are more rarely considered in theory that uses Gaussian fitness functions, perhaps because those functions are seen as general models of stabilizing selection, and therefore represent quite constrained models of nonlinear selection. The quadratic selection gradient matrix is given by

$$\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\beta}' - \mathbf{S} \tag{8}$$

(Chevin et al. 2015; note that this reference gives a mis-printed version of this relation). The expressions given here for selection gradients in Gaussian fitness models deviate somewhat from typical presentations in so far as we give multivariate expressions for all quantities.

It is possible to fit constrained functions (e.g., Gaussian) to empirical data in studies on natural selection, although this is rarely done (see Chevin et al. 2015 for an excellent example, linking theoretical quantities to selection gradients using a Gaussian model). More general exponential functions may be very useful for empirical studies of natural selection, particularly if they make less restrictive assumptions about the overall form of selection. If the coefficients of more general functions can be related analytically to selection gradients, they may help to better link theoretical and empirical evolutionary quantitative genetic studies. We sought to determine whether analytical relationships

could be found between selection gradients and the parameters of exponential functions with quadratic exponents, i.e.,

$$W(\mathbf{z}) = e^{a + \sum_i b_i z_i + \sum_i \frac{1}{2} g_i z_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} g_{ij} z_i z_j}. \tag{9}$$

Such functions are easily estimated via generalised linear models, using log link functions, and inherently benefit from the property of exponential functions to treat the response variable (expected fitness in this case) as a strictly positive quantity. The class of function we consider is also a generalization of the Gaussian fitness function (as in eq. 6), and is therefore linked to evolutionary quantitative genetic theory, but provides a more flexible model of nonlinear selection. Specifically, it need not be assumed that the function contains a maximum; when it does, relationships to theory that uses Gaussian functions may be invoked. As such, this more general approach, and its immediate link to statistical analysis with generalized linear models, may be very attractive to empiricists. We obtain analytical links between the regression parameters in equation 9 and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Our expressions coincide with known relations for Gaussian fitness functions (i.e., equations 7 and 8). The results are thus a particularly satisfying link between procedures that are likely to be adopted by empiricists, and the kinds of function that are used by theoreticians in evolutionary quantitative genetics. We also provide expressions for biological and statistical variance in selection gradients, given variance in the parameters of the regression in equation 9, and some links between exponential fitness functions and some other analyses about fitness components in use in evolutionary studies.

## Selection Gradients and Exponential Fitness Functions with Quadratic Arguments

Before detailing our results, a brief description of the factor of $\frac{1}{2}$ associated with the quadratic terms in equation 9, analogous to that surrounding a similar factor in Lande & Arnold's (1983) paper (see Stinchcombe et al. 2008), may prevent confusion. In order to obtain the correct values of the $g_i$ coefficients, the covariate values for quadratic terms should be squared and then halved. An alternative analysis is possible, where the squared covariate values are not halved, but the estimated coefficient estimates are doubled (analogous to procedures discussed by Stinchcombe et al. 2008). However, this alternative analysis leads to an additional, and potentially confusing, step in the calculation of standard errors and variance-covariance matrices of selection gradients in replicated studies (detailed in the appendix).

Define a vector $\mathbf{b} = (b_1, \ldots, b_k)'$ containing the coefficients of the linear terms in the exponent of the model in equation 9, and a matrix $\mathbf{g} = (g_{ij})$ containing the coefficients of the correspond-

ing quadratic form. We can then write the fitness function more conveniently in matrix form

$$W(\mathbf{z}) = e^{f(\mathbf{z})} \tag{10a}$$

$$f(\mathbf{z}) = a + \mathbf{b}'\mathbf{z} + \frac{1}{2}\mathbf{z}'\mathbf{g}\mathbf{z}. \tag{10b}$$

Let $\mathbf{d}$ be a vector of the expectations of the first order partial derivatives of $W(\mathbf{z})$ and let $\mathbf{H}$ be the matrix of expectations of the second order partial derivatives of $W(\mathbf{z})$. Thus the elements of $\mathbf{d}$ are $d_i = E[\frac{\partial W(\mathbf{z})}{\partial z_i}]$ and the elements of $\mathbf{H}$ are $H_{ij} = E[\frac{\partial^2 W(\mathbf{z})}{\partial z_i \partial z_j}]$. We can now rewrite the expressions for directional and quadratic selection gradients as

$$\boldsymbol{\beta} = \frac{\mathbf{d}}{E[W(\mathbf{z})]} \tag{11}$$

and

$$\boldsymbol{\gamma} = \frac{\mathbf{H}}{E[W(\mathbf{z})]}. \tag{12}$$

Differentiating equation 10 gives

$$\frac{\partial W(\mathbf{z})}{\partial \mathbf{z}'} = (\mathbf{b} + \mathbf{g}\mathbf{z})e^{f(\mathbf{z})}, \tag{13}$$

and

$$\frac{\partial^2 W(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}'} = \big(\mathbf{g} + (\mathbf{b} + \mathbf{g}\mathbf{z})(\mathbf{b} + \mathbf{g}\mathbf{z})'\big)e^{f(\mathbf{z})}. \tag{14}$$

Assume that the phenotype $\mathbf{z}$ is multivariate normal, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and denote its probability density by $p_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z})$. Provided $e^{f(\mathbf{z})}$ has a finite expectation, the function

$$K(\mathbf{z}) = \frac{e^{f(\mathbf{z})} p_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z})}{E[e^{f(\mathbf{z})}]} \tag{15}$$

is a probability density function giving the distribution of phenotypes after selection. Define the matrix $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} - \mathbf{g}$ and the vector $\mathbf{v} = \boldsymbol{\mu} + \boldsymbol{\Omega}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})$. We show in the Appendix that $\boldsymbol{\Omega}$ is symmetric. Provided it is also positive definite, it is a valid covariance matrix, and, by equation A8, $K(\mathbf{z}) \propto p_{\mathbf{v}, \boldsymbol{\Omega}}(\mathbf{z})$. As $K$ is a probability density function this implies

$$K(\mathbf{z}) = p_{\mathbf{v}, \boldsymbol{\Omega}}(\mathbf{z}). \tag{16}$$

Define $\mathbf{Q}^{-1} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma} = \mathbf{I}_k - \mathbf{g}\boldsymbol{\Sigma}$. Combining equations 11, 13 and 16 yields $\boldsymbol{\beta} = E[\mathbf{b} + \mathbf{g}\mathbf{z}]$, where the expectation is taken with respect to $K$. This is an expectation of a linear function of $\mathbf{z}$, and so

$$\boldsymbol{\beta} = \mathbf{b} + \mathbf{g}\mathbf{v} = (\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \mathbf{g}\boldsymbol{\Omega}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) = (\mathbf{I}_k + \mathbf{g}\boldsymbol{\Omega})(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})$$

$$= \mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}), \tag{17}$$

by use of equation A5.

Combining equations 12, 14, and 16 yields $\gamma = E[\mathbf{g} + (\mathbf{b} + \mathbf{gz})(\mathbf{b} + \mathbf{gz})']$, where the expectation is taken with respect to $K$. Hence

$$
\begin{aligned}
\gamma &= \mathbf{g} + \mathrm{VAR}(\mathbf{b} + \mathbf{gz}) + [E(\mathbf{b} + \mathbf{gz})][E(\mathbf{b} + \mathbf{gz})]' \\
&= \mathbf{g} + \mathbf{g\Omega g}' + \boldsymbol{\beta\beta}' \\
&= \boldsymbol{\beta\beta}' + (\mathbf{I}_k + \mathbf{g\Omega})\mathbf{g} \\
&= \boldsymbol{\beta\beta}' + \mathbf{Qg},
\end{aligned} \tag{18}
$$

where we have noted that $\mathbf{g}$ is symmetric and used equation A5.

In univariate analyses, the matrix machinery necessary for implementing the general formulae in equations 17 and 18 can be avoided. If the fitness function is $W(z) = e^{a + bz + \frac{1}{2}gz^2}$, and $z$ has a mean of $\mu$ and a variance of $\sigma^2$, then $\beta = \frac{b + g\mu}{1 - g\sigma^2}$ and $\gamma = \frac{(b + g\mu)^2 + g(1 - g\sigma^2)}{(1 - g\sigma^2)^2}$. These expressions will hold for any univariate analysis, and can be applied to get mean-standardized, variance-standardized, and unstandardized selection gradients, when appropriate values of $\mu$ and $\sigma^2$ are used, and applied to log-quadratic models of $W(z)$ where the phenotypic records have been correspondingly standardized. For the common case where the trait is mean-centred and (unit) variance standardized, the expressions simplify further to $\beta = \frac{b}{1 - g}$ and $\gamma = \frac{b^2 + g(1 - g)}{(1 - g)^2}$.

The notation we have used is designed to relate the parameters one may estimate directly in a regression analysis to selection gradients. This should facilitate the application of log-linear models of fitness functions in empirical models of selection. Interestingly, the expressions we obtain, while more general than Gaussian functions (they do not constrain selection to have a stabilising form), coincide with those for Gaussian selection. This equivalence is demonstrated in the appendix. We have thus shown that these expressions are more general and empirically useful than has previously been known. The requirements for obtaining selection gradients from Gaussian and the more general exponential functions with quadratic arguments are slightly different. For Gaussian functions, $\omega$ (as in equation 6) must be positive definite. Equations 17 and 18 require that $\boldsymbol{\Omega}$, which depends on both $\mathbf{g}$ and $\boldsymbol{\Sigma}$, is positive definite. In univariate analyses, this condition reduces to $g < \frac{1}{\sigma^2}$, implying that the fitness function should not curve upward too sharply within the range of observed phenotype.

The main application of the expressions given here to obtain selection gradient estimates from log-linear or log-quadratic arguments is generalized linear model (GLM) regression analyses with log link functions. With appropriate specification of linear, quadratic, and correlational terms, GLMs with any response variable distribution should yield selection gradient estimates, providing that they specify a log link function. This will include, for example, binomial and negative binomial GLM models, with arguments specified as in equation 4 or 9. The equations also apply directly to log-link models with additive overdispersion (appendix section). Additionally under certain conditions, sev-

eral other analyses commonly used to assess the dependence of fitness, or fitness components, on quantities such as phenotypic traits can yield log-linear or log-quadratic models of trait-fitness relationships, and can thus be used with the expressions given in this section for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. These include special cases of parentage analysis, capture-mark-recapture analysis, and survival analysis. The conditions under which these methods can yield selection gradient estimates are elaborated in appendix sections.

## Biological Variation and Statistical Uncertainty

The expressions for selection gradients, given the parameters of a log-quadratic fitness function (equations 17 and 18) give the selection gradients conditional on the estimated values of $\mathbf{b}$ and $\mathbf{g}$. However, we often make inference about selection, not merely to quantify selection at a given place and time, but rather to ask larger questions, for example, about how selection varies. For exponential models with a linear argument (equation 4), the variance in $\mathbf{b}$ among, for example, cohorts or populations, may be estimated by random-regression analysis, and would represent the variance in directional selection gradients, since $\boldsymbol{\beta} = \mathbf{b}$. However, since the $b$ and $g$ coefficients in an exponential model of fitness with a quadratic argument (equation 9) are not themselves selection gradients, their variances and covariances are not the variances and covariances of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Moreover in empirical studies of natural selection, $\mathbf{b}$ and $\mathbf{g}$ will not typically be known quantities, but rather will be estimates with error.

Whether variances and covariances of the elements of $\mathbf{b}$ and $\mathbf{g}$ are of biological interest (e.g., variation among temporal replicates of a selection analysis), or are statistical (e.g., sampling variance), the corresponding biological or sampling variances of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can potentially be obtained by integrating approximations, bootstrapping, and/or Monte Carlo methods. In particular, approximation of variances in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by a first-order Taylor approximation (the "delta method"; Lynch & Walsh 1998) may generally be pragmatic. Formulae for approximate biological or statistical variances of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given covariances of $\mathbf{b}$ and $\mathbf{g}$ by this method are given in the appendix. For univariate analysis, with phenotype standardised to $\mu = 0$ and $\sigma^2 = 1$, the approximate variances $\beta$ and $\gamma$ are given by

$$
\mathrm{Var}[\beta] \approx \frac{Var[b]}{(1 - g)^2} + \frac{b^2 Var[g]}{(1 - g)^4} + \frac{2b Cov[b, g]}{(1 - g)^3}, \tag{19}
$$

and

$$
\mathrm{Var}[\gamma] \approx \frac{4b^2 Var[b]}{(1 - g)^4} + \frac{(1 + 2b^2 - g)^2 Var[g]}{(1 - g)^6} + \frac{4b(1 + 2b^2 - g)Cov[b, g]}{(1 - g)^5}, \tag{20}
$$

where $Var[b]$ and $Var[g]$ represent the variances of $b$ and $g$ terms and $Cov[b, g]$ is the covariance of the $b$ and $g$ terms. These variances and covariances may represent real variation in trait-fitness relationships (e.g., variation in time or space), or they may represent statistical uncertainty in parameter values. If equations 19 and 20 are used to represent statistical uncertainty in estimates $\hat{\beta}$ and $\hat{\gamma}$, $Var[b]$ would be replaced by $Var[\hat{b}]$ (the sampling variance, or squared standard error of the estimate of $b$), etc. The sampling covariance of $b$ and $g$ (i.e., $Cov[\hat{b}, \hat{g}]$) is not routinely reported by most statistical software packages, but can generally be obtained (the use of the sampling covariance of $\hat{b}$ and $\hat{g}$ terms in a GLM model in R is demonstrated in the supplemental materials).

## STATISTICAL UNCERTAINTY: SIMULATION

We performed a small simulation study to assess the extent of any bias in the estimators $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and the adequacy of the first-order approximation of their standard errors. We simulated univariate directional selection, with values of $b$ between -0.5 and 0.5, and with $g = -1, 0$, and 0.4. Because $\beta$ and $\gamma$ are nonlinear functions of $g$, it is not possible to simultaneously investigate ranges of parameter values with regular intervals of values of both $g$ and selection gradients. These values of $g$ represent a compromise between investigating a regular range of $g$ and $\gamma$. We used a (log) intercept of the fitness function of $a = 0$. We simulated a sample size of 200 individuals. This sample size reflects a very modest-sized study with respect to precision in inference of nonlinear selection, and is therefore a useful scenario in which to judge performance of different methods for calculating standard errors. Fitness was simulated as a Poisson variable with expectations defined by the ranges of values of $b$ and $g$, and with phenotypes sampled from a standard normal distribution.

First, we analyzed each simulated dataset using the OLS regression described by Lande & Arnold (1983), i.e., $w_i = \mu + \beta z_i + \frac{1}{2} \gamma z_i^2 + e_i$, using the R function `lm()`. For the OLS regressions, we calculated standard errors assuming normality using the standard method implemented in the R function `summary.lm()`, and by case-bootstrapping, by generating 1000 bootstrapped datasets by sampling with replacement, running the OLS regression analysis, and calculating the standard deviation of the bootstrapped selection gradient estimates. Second, we fitted a Poisson GLM with linear and quadratic terms, using the R function `glm()`. We then calculated conditional selection gradient estimates using equations 17 and 18. We obtained standard errors by using a first-order Taylor series approximation (the "delta method"; Lynch & Walsh 1998, appendix A1). For each method of obtaining estimates and standard errors, we calculated the standard deviation of replicate simulated estimates. We could thus evaluate the performance of different methods of obtaining standard errors by their ability to reflect this sampling standard deviation. We also calculated mean absolute errors for both estimators of $\beta$ and $\gamma$ for all scenarios. Every simulation scenario and associated analysis of selection gradients was repeated 1000 times.

Selection gradient estimates obtained by all three methods were essentially unbiased (Figure 1a,d,g,j,m,p), except for small biases that occurred when the fitness function was very curved. Thus, GLM-derived values of selection gradients, conditional on estimated values of $b$ and $g$, performed very well as estimators of $\beta$ and $\gamma$ in our simulations. Similarly, first-order approximations of standard errors of the GLM-derived estimates of $\beta$ and $\gamma$ closely reflected the simulated standard deviations of the estimators (Figure 1b,e,h,k,n,q). All methods for obtaining standard errors performed well for estimates of $\beta$ in the pure log-linear selection simulations (Figure 1h,k). OLS standard errors performed reasonably well under most simulation scenarios, except when $g$ was positive (Figure 1n,q); across all scenarios bootstrap standard errors of the OLS estimators outperformed OLS standard errors produced using the standard formula. Mean squared error of the GLM estimators was always smaller than that of the OLS estimators of $\beta$ and $\gamma$. This is unsurprising, as the simulation scheme corresponded closely to the GLM model. These results demonstrate the usefulness of the conditional values of $\beta$ and $\gamma$ as estimators, and show that gains in precision and accuracy can be obtained when glm models of fitness functions fit the data well.

It remains plausible that the OLS estimators motivated by Lande & Arnold's (1983) work could outperform glm-based analyses in some scenarios. In particular, the OLS estimators will rarely be misleading (Figure 1), whereas the risk of bias or misleading standard errors in workflows involving GLMs resulting from, for example, mis-specification of error structures, is not known. What the theory and limited simulations presented here demonstrate is that (a) given a well-specified log-link GLM, selection gradients can be obtained and (b) first-order approximations to the standard errors of these GLM-based selection gradients seem to perform adequately.

## VARIATION IN SELECTION: EMPIRICAL EXAMPLE

In order to explore the behavior of estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ derived from log-quadratic fitness models, we conducted a small study of selection and variation in selection of birth mass in female Soay sheep (*Ovis aries*), from a long-term study on St Kilda, in the Outer Hebrides, Scotland (Clutton-Brock & Pemberton 2004). Briefly, the data comprise records of mass taken within 5 days of birth, and subsequent lifetime breeding success (i.e., the total number of offspring produced, regardless of the fates of those offspring) of females from cohorts born between 1985 and 2006 (inclusive, except for 2001, when data collection was suspended due to precautions associated with an outbreak of foot-and-mouth disease). Fitness data were collected between 1985
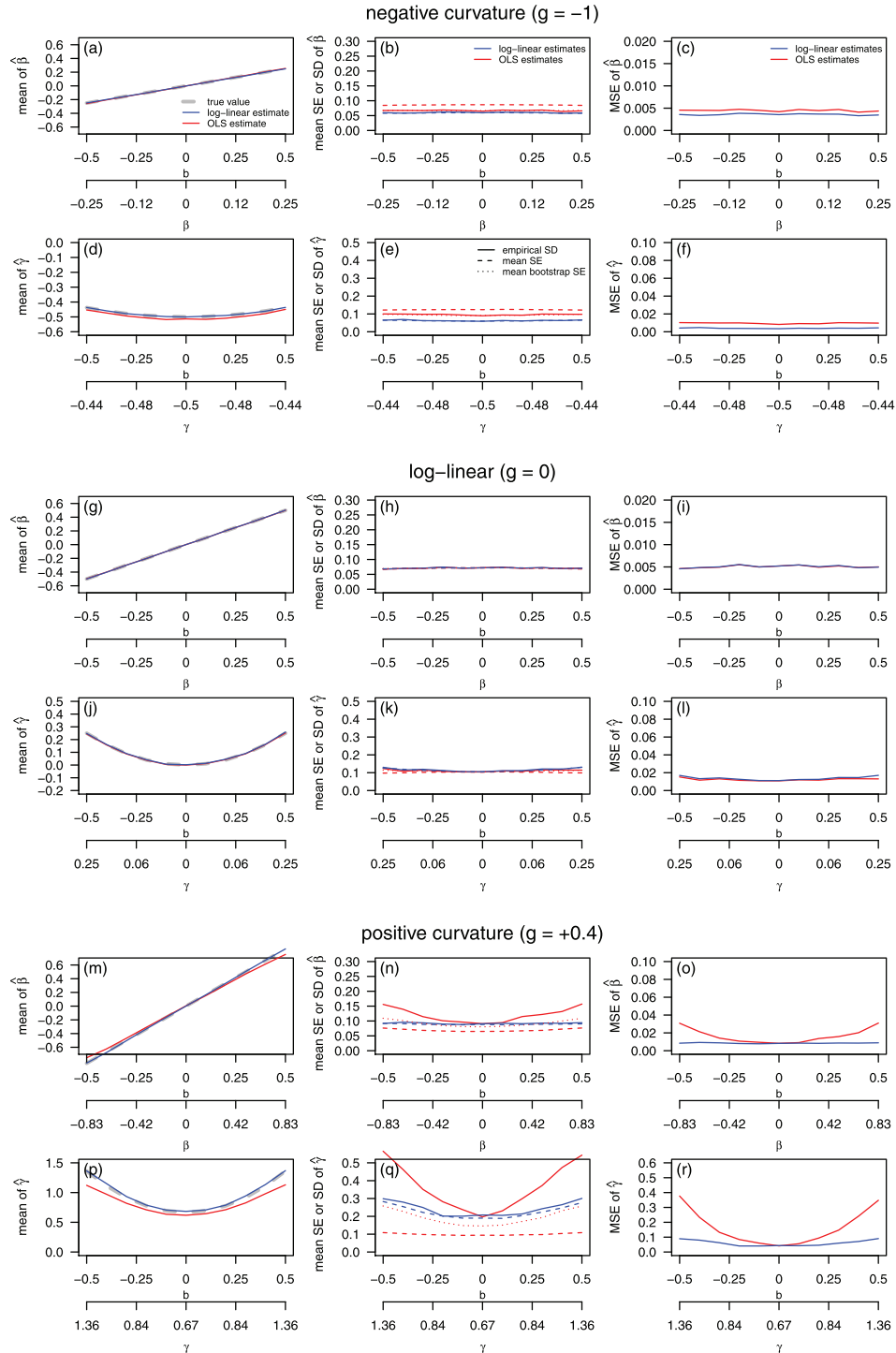
**Figure 1.** Simulation results for the performance of Lande & Arnold's (1983) least squares-based (OLS) estimators (red lines), and log-quadratic (GLM) estimators (blue lines), of directional and quadratic selection gradients. The first column shows bias in estimates of β and γ, where departure from the grey line (the simulated truth) indicates bias. The middle column shows the performance of OLS standard errors (red dashed lines), bootstrap standard errors (red dotted lines), and first-order approximations (blue dashed lines) of the standard errors of the GLM estimators. Ideally, all values of estimated mean standard errors would fall on the simulated standard deviation of their associated estimators, shown as solid lines. The right column shows the mean squared errors of the OLS and GLM estimators. Note that the *y*-axis scale for the MSE in plot o differs from that in plots (c) and (i), that the scale in plot (r) differs from that in plots (f) and (l), and that the scales in plots (d), (j), and (p) differ.

and 2016. A small number of ewes from later cohorts (i.e., 2005 and 2006) will have still been alive in 2016, and their fitness will therefore be slightly underestimated. Birth masses were corrected for growth in the first 5 days prior to other analyses by fitting a linear model of mass as a function of age, and then correcting mass for growth.

We fitted a single model as our basis for estimating the variance in selection among cohorts. We fitted a generalized linear random regression mixed model describing the dependence of lifetime breeding success, $W_{ij}$, of individual $i$ from cohort $j$, as a function of their birth mass $z_{ij}$, according to

$$W_{ij} = P\big(E[W]_{ij}\big), \tag{21a}$$

$$log\big(E[W]_{ij}\big) = a_j + \bar{b}z_{ij} + \frac{1}{2}\bar{g}z_{ij}^2 + \delta_{b,j}z_{ij}^2 + \frac{1}{2}\delta_{g,j}z_{ij} + e_{ij}, \tag{21b}$$

where $P(\lambda)$ denotes samples from a Poisson distribution with expectation $\lambda$, $a_j$ are cohort-specific intercepts, and $\bar{b}$ and $\bar{g}$ are overall log-scale slopes and curvatures. $\delta_{b,j}$ and $\delta_{g,j}$ are cohort-specific random effects for slopes and curvature terms, assumed to be distributed according to

$$\begin{bmatrix} \delta_b \\ \delta_g \end{bmatrix}_j \sim N(\mathbf{0}, \Sigma_\delta),$$

and elements of the covariance matrix $\Sigma_\delta = \begin{bmatrix} \sigma_{\delta_b}^2 & \sigma_{\delta_b,\delta_g} \\ \sigma_{\delta_b,\delta_g} & \sigma_{\delta_g}^2 \end{bmatrix}$ are estimated. $e_{ij}$ are residuals, with zero means and cohort-specific variances. We implemented the model defined by equations 21 as a mixed model, and collected MCMC samples, using `MCMCglmm` (Hadfield 2010). We used diffuse normal priors on the fixed effects, an inverse-Wishart prior for $\Sigma_\delta$, and inverse-gamma priors on the cohort-specific log-scale overdispersion variances, $\sigma_e^2$. We used informative priors on the cohort-specific overdispersion variances, $\sigma_e^2$, in order to help the model to predict cohort-specific mean fitnesses that agreed well with the observed data. The specific model is fully detailed in the code supplement.

Selection is predominantly directional, favoring larger birth masses (Table 1). Quadratic selection is near zero on average, but not because there is generally a lack of curvature of the fitness function. The combination of positive values of $b$ and negative values of $g$ (from fitting equation 21; Table 1 a) implies that the fitness function is often upwardly curved over much of the distribution of birth mass, but becomes downwardly curved in the region of large birth masses (Figure 2), such that the curvature, averaged over the distribution of phenotype, tends to be near zero. Using the system given in the appendix for calculating variances in $\beta$ and $\gamma$ (with univariate cases given in equations 19 and 20), we inferred that there is very substantial variation in both direc-

**Table 1.** Parameters of the random regression mixed model (a) used to characterize variation in selection of lamb body mass in Soay sheep (*Ovis aries*). Part (b) gives estimates of the means, variances, and covariances of selection gradients, derived from the model reported in part (a). For ease of interpretation, estimates of variation in β and γ are reported as standard deviations, and the relationship between β and γ is reported as a correlation.

| Parameter | Estimate and 95% CI |
|---|---|
| (a) Mixed model parameters | |
| $b$ | 1.07 ( 0.83 – 1.35) |
| $g$ | −0.72 (−1.08 – -0.33) |
| $\sigma_d^2$ | 0.25 (0.05 – 0.50) |
| $\sigma_e^2$ | 0.47 (0.07 – 1.02) |
| $\sigma_{d,e}$ | −0.26 (−0.56 – 0.00) |
| (b) Derived parameters of the distribution of selection gradients | |
| $\mu_\beta$ | 0.62 ( 0.52 – 0.74) |
| $\mu_\gamma$ | 0.02 (-0.21 – 0.21) |
| $\sigma_\beta$ | 0.18 ( 0.04 – 0.34) |
| $\sigma_\gamma$ | 0.35 ( 0.06 – 0.71) |
| $\rho_{\beta,\gamma}$ | 0.64 (-0.18 – 1.00) |

tional and quadratic selection. While β varies greatly ($\hat{\sigma}_\beta = 0.18$ (0.04 – 0.34); Table 1 b), because selection is so strongly directional, this does not lead to fluctuations in the direction of selection. The fact that the average quadratic selection gradient is near zero does not imply that nondirectional selection is absent. Rather, in contrast to the situation for directional selection, the estimate of $\hat{\sigma}_\gamma = 0.35$ (0.06 – 0.71) (Table 1 b), given that $\hat{\mu}_\gamma = 0.02$ (–0.21 – 0.21), suggests quadratic selection varies substantially, sometimes taking positive values, and sometimes being negative.

Annual (log) quadratic functions from the random regression mixed model (equation 21) correspond reasonably closely to quadratic OLS estimates, for most cohorts (Figure 2). There are some systematic differences, arising primarily from the fact that the two models have different functional forms; the former is a quadratic function, while the latter is a quadratic function on the log scale, or a Gaussian function when the parameter $g$ is negative, as it often is (Table 1 a).

## Conclusion

We have provided analytical expressions for selection gradients, given the parameters of exponential fitness models with quadratic arguments. These functions can be applied in conjunction with a range of generalised linear model approaches, specific situations in capture-mark-recapture, survival analysis, and parentage analysis, and relate empirical selection gradients directly to types of fitness functions used in theoretical studies. The general
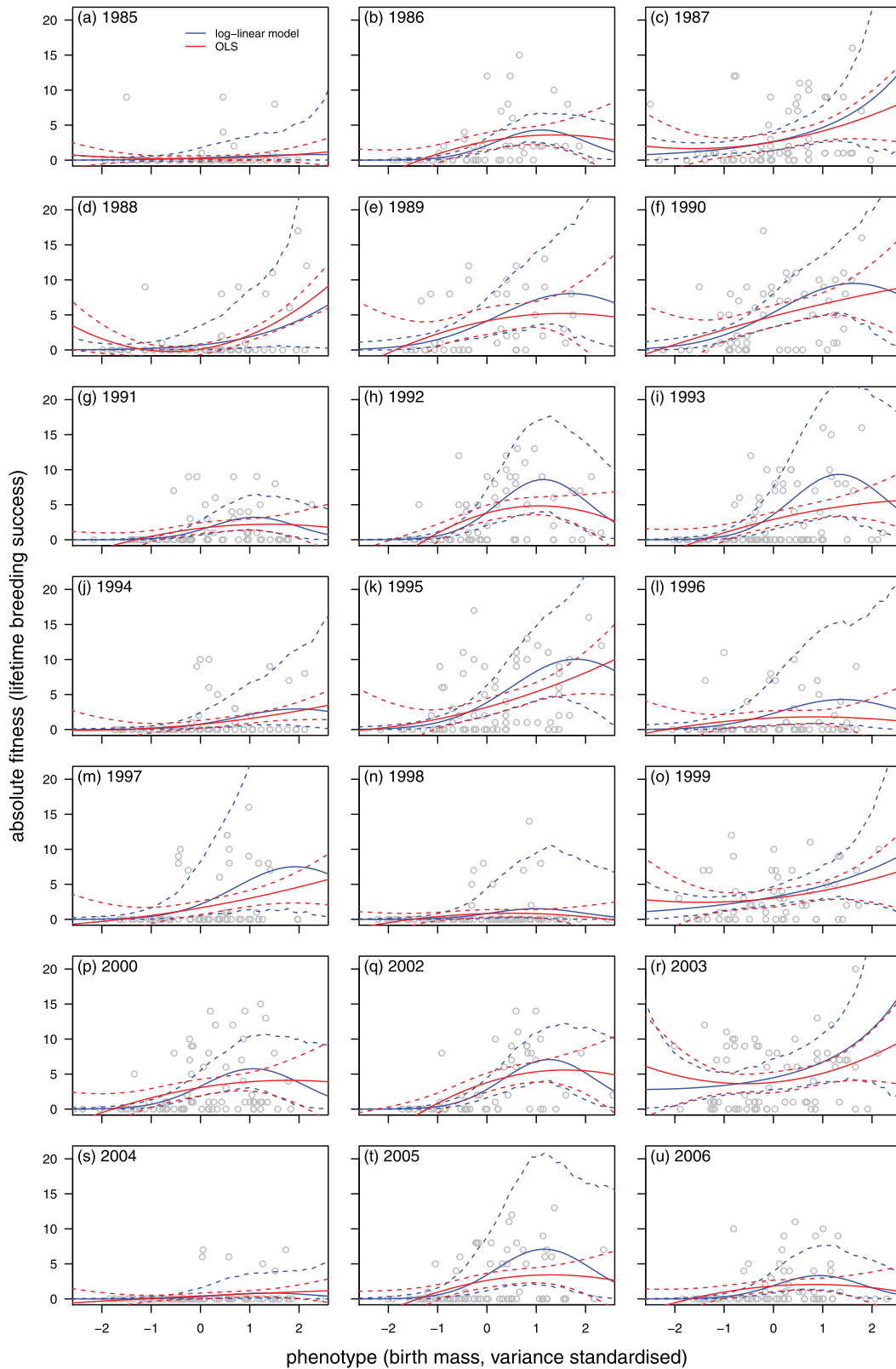
**Figure 2.** Estimates of functions relating birth mass to lifetime breeding success in 21 cohorts of Soay sheep (*Ovis aries*) on St Kilda from 1985 to 2006; note that inference of selection in 2001 is not possible because very little phenotypic data were collected on account of the foot and mouth outbreak in the UK that year. Estimated fitness functions are given with 95% confidence intervals (OLS) or credible intervals (Bayesian generalized linear mixed model or GLMM) intervals of the predictions. OLS estimates are generated independently for each cohort; GLMM estimates depict cohort-specific predictions from the random regression mixed model described in equation 21.

relationship of selection gradients to the coefficients of log-linear and log-quadratic models, and in particular, various ways of estimating these using generalized linear models, are probably the most generally useful feature of our results. In empirical applications, our preliminary simulation results indicate that, given an appropriate model of a log-scale fitness function, inference using log-linear and log-quadratic models may be very robust, and could provide more reliable statements about uncertainty (i.e., reasonable standard errors) than the main methods used to date. It should be noted, however, that OLS methods proved to be highly robust in our simulations with Poisson fitness residuals (see also McGee, submitted), except in the presence of very strong nonlinear selection. Furthermore, the relationships given here between log-quadratic fitness functions and selection gradients could lead to better integration between empirical and theoretical strategies for modelling selection.

## AUTHOR CONTRIBUTIONS

## ACKNOWLEDGMENTS

## LITERATURE CITED

Aitchison, J. & Brown, J. (1957) The Lognormal Distribution, with special reference to its uses in economics. Cambridge: Cambridge University Press.

Chevin, L.M. & Haller, B.C. (2014) The temporal distribution of directional gradients under selection for an optimum. *Evolution*, 68, 3381–3394.

Chevin, L.M. & Hospital, F. (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180, 1645–1660.

Chevin, L.M., Visser, M.E. & Tufto, J. (2015) Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. *Evolution*, 69, 2319–2332.

Clutton-Brock, T.H. & Pemberton, J.M. (2004) Soay sheep dynamics and selection in an island population. Cambridge: Cambridge University Press.

Cox, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34, 187–220.

Endler, J.A. (1986) Natural selection in the wild. Princeton University Press.

Falconer, D.S. (1960) Introduction to Quantitative Genetics. Oliver and Boyd.

Gomulkiesicz, R. & Houle, D. (2009) Demographic and genetic constraints on evolution. *The American Naturalist*, 174, E218–E229.

Hadfield, J. (2010) MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22.

Hadfield, J.D., Richardson, D.S. & Burke, T. (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, 15, 3715–3731.

Janzen, F.J. & Stern, H.S. (1998) Logistic regression for empirical studies of multivariate selection. *Evolution*, 52,1564–1571.

Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Vignieri, C., Berrigan, D., Hill, E., Hoang, A., Gilbert, P. & Beerli, P. (2001) The strength of phenotypic selection in natural populations. *The American Naturalist*, 157, 245–261.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30, 314–334.

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

Lande, R. (1983) The response to selection on major and minor mutations affecting a metrical trait. *Heredity*, 50, 47–55.

Lande, R. & Arnold, S.J. (1983) The measurement of selection on correlated characters. *Evolution*, 37, 1210–1226.

Lebreton, J.D., Burnham, K.P., Colbert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 62, 67–118.

Lynch, M. & Walsh, B. (1998) Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer.

Manly, B.F.J. (1985) The statistics of natural selection. New York: Chapman and Hall.

Morrissey, M.B. & Sakrejda, K. (2013) Unification of regression-based approaches to the analysis of natural selection. *Evolution*, 67, 2094–2100.

Schluter, D. (1988) Estimating the form of natural selection on a quantitative trait. *Evolution*, 42, 849–861.

Shaw, R.G. & Geyer, C.J. (2010) Inferring fitness landscapes. *Evolution*, 64, 2510–2520.

Smouse, P.E., Meagher, T.R. & Kobak, C.J. (1999) Parentage analysis in *Chamaelirium luteum* (L.) gray (Liliaceae): why do some males have higher reproductive contributions? *Journal of Evolutionary Biology*, 12, 1069–1077.

Stinchcombe, J.R., Agrawal, A.F., Hohenlohe, P.A., Arnold, S.J. & Blows, M.W. (2008) Estimating nonlinear selection gradients using quadratic regression coefficients: Double or nothing? *Evolution*, 62, 2435–2440.

Waller, J. & Svensson, E. (2016) The measurement of selection when detection is imperfect: how good are naïve methods? Methods in Ecology and Evolution, 7, 538–548.

Walsh, B. & Lynch, M. (2018) Evolution and selection of quantitative traits. Oxford, UK: Oxford University Press.

Weldon, W.F.R. (1901) A first study of natural selection in *Clausilia italica* (von martens). *Biometrika*, 1, 109–124.

## *Appendix*

### GAUSSIAN FITNESS FUNCTIONS AS SPECIAL CASES OF LOG NORMAL FITNESS FUNCTIONS WITH QUADRATIC ARGUMENTS

Suppose that the fitness function $W(\mathbf{z})$ is proportional to a Gaussian density function for $\mathbf{z}$ with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\omega}$. The exponent of the corresponding exponential function is then given by

$$-\frac{1}{2}(\mathbf{z}-\boldsymbol{\theta})'\boldsymbol{\omega}^{-1}(\mathbf{z}-\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{\theta}'\boldsymbol{\omega}^{-1}\boldsymbol{\theta} + \boldsymbol{\theta}'\boldsymbol{\omega}^{-1}\mathbf{z} - \frac{1}{2}\mathbf{z}'\boldsymbol{\omega}^{-1}\mathbf{z} = a + \mathbf{b}'\mathbf{z} + \frac{1}{2}\mathbf{z}'g\mathbf{z},$$

where $a = -\frac{1}{2}\boldsymbol{\theta}'\boldsymbol{\omega}^{-1}\boldsymbol{\theta}$, $\mathbf{b}' = \boldsymbol{\theta}'\boldsymbol{\omega}^{-1}$ and $g = -\boldsymbol{\omega}^{-1}$. Let $\mathbf{z}$ have a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Setting $\mathbf{S} = (\boldsymbol{\omega} + \boldsymbol{\Sigma})^{-1}$, the selection gradient vector $\boldsymbol{\beta}$ given by equation 7 is

$$-\mathbf{S}(\boldsymbol{\mu} - \boldsymbol{\theta}) = (\boldsymbol{\omega} + \boldsymbol{\Sigma})^{-1}(\boldsymbol{\omega}\mathbf{b} - \boldsymbol{\mu}) = (-\mathbf{g}^{-1} + \boldsymbol{\Sigma})^{-1}(-\mathbf{g}^{-1}\mathbf{b} - \boldsymbol{\mu})$$

$$= (\mathbf{g}^{-1} - \boldsymbol{\Sigma})^{-1}\mathbf{g}^{-1}\mathbf{g}(\mathbf{g}^{-1}\mathbf{b} + \boldsymbol{\mu})$$

$$= (\mathbf{I}_k - \mathbf{g}\boldsymbol{\Sigma})^{-1}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})$$

$$= \mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}), \qquad (A1)$$

where $\mathbf{Q}$ is defined following equation 16. This shows that the result for the selection gradient using Gaussian fitness functions is a special case of our result in equation 17.

Comparison of the pre-multipliers of $\boldsymbol{\mu}$ in A1 indicates that $-\mathbf{S} = \mathbf{Q}\mathbf{g}$. It follows that equation 8, giving the result for quadratic selection gradients using Gaussian fitness functions, is a special case of our result in equation 18.

## SAMPLING VARIANCES OF SELECTION GRADIENT ESTIMATES

Denote a vector containing all unique elements of $\boldsymbol{\gamma}$ by $\tilde{\boldsymbol{\gamma}}$. The following assumes that $\tilde{\boldsymbol{\gamma}}$ is composed by vertically stacking the columns of the diagonal and sub-diagonal elements of $\boldsymbol{\gamma}$. For example, in an analysis with three traits, $\tilde{\boldsymbol{\gamma}} = [\gamma_{1,1}, \gamma_{2,1}, \gamma_{3,1}, \gamma_{2,2}, \gamma_{3,2}, \gamma_{3,3}]'$. Let $\mathbf{v}()$ denote the function mapping the distinct elements of a symmetric matrix $\mathbf{r}$ onto the column vector $\tilde{\mathbf{r}}$.

The first-order approximation to the sampling covariance matrix of the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is then given by $\mathbf{J}\tilde{\boldsymbol{\Sigma}}\mathbf{J}'$, where $\tilde{\boldsymbol{\Sigma}}$ is the sampling covariance matrix of a vector containing the elements of $\mathbf{b}$ and $\tilde{\mathbf{g}}$, where the latter is a column vector containing the distinct elements of $\mathbf{g}$ arranged according to the same scheme that defines $\tilde{\boldsymbol{\gamma}}$. $\mathbf{J}$ is the Jacobian, or gradient matrix of first order partial derivatives, of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\gamma}}$ with respect to $\mathbf{b}$ and $\tilde{\mathbf{g}}$, i.e.,

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}} & \frac{\partial \boldsymbol{\beta}}{\partial \tilde{\mathbf{g}}} \\ \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \mathbf{b}} & \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \tilde{\mathbf{g}}} \end{bmatrix},$$

evaluated at the estimated values of $\mathbf{b}$ and $\mathbf{g}$.

Note that some users may prefer to fit the model 9 with $g_{ii}$ replaced by $2g_i$, say. The formulae for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are readily re-expressed in terms of these variables by making this substitution. If $\boldsymbol{\Sigma}_1$ denotes the covariance matrix obtained when fitting this revised model, the required covariance matrix $\tilde{\boldsymbol{\Sigma}}$ can be calculated using $\tilde{\boldsymbol{\Sigma}} = \mathbf{D}\boldsymbol{\Sigma}_1\mathbf{D}'$, where $\mathbf{D}$ is a diagonal matrix with all the diagonal elements equal to one, apart from those corresponding to the variables $g_{ii}$ which equal 2.

The four submatrices of $\mathbf{J}$ can be treated separately. Noting that $\boldsymbol{\beta} = \mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})$ (equation 17),

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}} = \mathbf{Q}. \qquad (A2)$$

Let $s = \frac{1}{2}k(k+1)$, where $k$ is the number of traits in the analysis, and let $\mathbf{e}_1, \ldots, \mathbf{e}_s$ be the standard basis for an $s$ dimensional space (i.e., $\mathbf{e}_1 = [1, 0, \ldots, 0]'$, etc.). Define an indicator matrix $\mathbf{C}_m = \mathbf{C}^{(i,j)}$ where $\mathbf{C}^{(i,j)}$ is a $k$ by $k$ matrix in which

$$[\mathbf{C}^{(i,j)}]_{xy} = \begin{cases} 1, & (x, y) = (i, j) \text{ or } (j, i); \\ 0, & \text{otherwise.} \end{cases}$$

Using the standard expression for the derivative of the inverse of a matrix with respect to a scalar, we can obtain $\frac{\partial \boldsymbol{\beta}}{\partial \tilde{\mathbf{g}}}$, i.e., the upper-right sub-matrix of $\mathbf{J}$.

$$\boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) \quad \Rightarrow \frac{\partial \boldsymbol{\beta}}{\partial \tilde{g}_m} = \frac{\partial \boldsymbol{\beta}}{\partial g_{ij}} = -\boldsymbol{\Psi}^{-1}\left[\frac{\partial \boldsymbol{\Psi}}{\partial g_{ij}}\right]\boldsymbol{\Psi}^{-1}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \boldsymbol{\Psi}^{-1}\left[\frac{\partial(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})}{\partial g_{ij}}\right]$$

$$= -\mathbf{Q}\left[\frac{\partial \mathbf{I}_k - \mathbf{g}\boldsymbol{\Sigma}}{\partial g_{ij}}\right]\mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right]\boldsymbol{\mu}$$

$$= \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right][\boldsymbol{\Sigma}\mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})] + \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right]\boldsymbol{\mu}$$

$$= \mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu}) = \mathbf{Q}\mathbf{C}_m(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})$$

$$\Rightarrow \frac{\partial \boldsymbol{\beta}}{\partial \tilde{\mathbf{g}}} = \sum_{m=1}^{s} \frac{\partial \boldsymbol{\beta}}{\partial \tilde{g}_m}\mathbf{e}'_m = \mathbf{Q}\sum_{m=1}^{s}\mathbf{C}_m(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})\mathbf{e}'_m \qquad (A3)$$

Let $\mathbf{Q}_{[u]}$ denote the $u^{th}$ column of $\mathbf{Q}$. Using the previous relation $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}} = \mathbf{Q}$, we can obtain $\frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \mathbf{b}}$, i.e., the lower-left sub-matrix of $\mathbf{J}$.

$$\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{Q}\mathbf{g} \quad \Rightarrow \frac{\partial \boldsymbol{\gamma}}{\partial b_u} = \boldsymbol{\beta}\left(\frac{\partial \boldsymbol{\beta}}{\partial b_u}\right)' + \left(\frac{\partial \boldsymbol{\beta}}{\partial b_u}\right)\boldsymbol{\beta}' = \boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}'$$

$$\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial b_u} = \mathbf{v}(\boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}')$$

$$\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \mathbf{b}} = \sum_{u=1}^{k} \mathbf{v}(\boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}')\mathbf{e}'_u \qquad (A4)$$

Let $\mathbf{M}^{(m)} = \mathbf{Q}\mathbf{C}_m(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})\boldsymbol{\beta}'$. Note that $\mathbf{Q}^{-1} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}$ implies $\boldsymbol{\Omega} = \boldsymbol{\Sigma}\mathbf{Q}$. Moreover $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} - \mathbf{g}$ implies firstly that

$$\mathbf{I}_k + \mathbf{g}\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} = \mathbf{Q} \qquad (A5)$$

and secondly that $\boldsymbol{\Omega}$ is symmetric, since $\boldsymbol{\Sigma}$ and $\mathbf{g}$ are both symmetric. It follows that

$$\mathbf{Q}' = \mathbf{I}_k + (\mathbf{g}\boldsymbol{\Omega})' = \mathbf{I}_k + \boldsymbol{\Omega}\mathbf{g}. \qquad (A6)$$

The lower-right sub-matrix of $\mathbf{J}$ can then be derived.

$$\frac{\partial \boldsymbol{\gamma}}{\partial g_{ij}} = \left[\frac{\partial \boldsymbol{\beta}}{\partial g_{ij}}\right]\boldsymbol{\beta}' + \boldsymbol{\beta}\left[\frac{\partial \boldsymbol{\beta}}{\partial g_{ij}}\right]' + \mathbf{Q}\mathbf{C}^{(ij)} + \mathbf{Q}\mathbf{C}^{(ij)}\boldsymbol{\Sigma}\mathbf{Q}\mathbf{g}$$

$$= [\mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})]\boldsymbol{\beta}' + \boldsymbol{\beta}[\mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})]' + \mathbf{Q}\mathbf{C}^{(ij)} + \mathbf{Q}\mathbf{C}^{(ij)}\boldsymbol{\Omega}\mathbf{g}$$

$$\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial g_{ij}} = \mathbf{v}[\mathbf{M}^{(m)} + (\mathbf{M}^{(m)})' + \mathbf{Q}\mathbf{C}_m(\mathbf{I}_k + \boldsymbol{\Omega}\mathbf{g})]$$

$$\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \tilde{\mathbf{g}}} = \sum_{m=1}^{s} \mathbf{v}[\mathbf{M}^{(m)} + (\mathbf{M}^{(m)})' + \mathbf{Q}\mathbf{C}_m\mathbf{Q}']\mathbf{e}'_m,$$

$$(A7)$$

by use of equation A6.

Finally note that equations A5 and A6 are also relevant to the derivation of formula 16. By definition, $f(\mathbf{z}) = a + \mathbf{z}'\mathbf{b} + \frac{1}{2}\mathbf{z}'\mathbf{g}\mathbf{z}$, and we have $\log[p_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{z})] = -\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z} + \mathbf{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \alpha$, where $\alpha$

does not depend on $\mathbf{z}$. Thus, if $\alpha' = \alpha + a$, it follows that, as a function of $\mathbf{z}$,

$$f(\mathbf{z}) + \log[p_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})] = -\tfrac{1}{2}\mathbf{z}'(\boldsymbol{\Sigma}^{-1} - \mathbf{g})\mathbf{z} + \mathbf{z}'(\mathbf{b} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}) + \alpha'$$

$$= -\tfrac{1}{2}\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} + \mathbf{z}'\boldsymbol{\Omega}^{-1}\big[\boldsymbol{\Omega}(\mathbf{b} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})\big] + \alpha',$$

Now, by A5 and A6, we have $\boldsymbol{\Omega}(\mathbf{b} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}) = \boldsymbol{\Omega}\mathbf{b} + (\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega})'\boldsymbol{\mu} = \boldsymbol{\Omega}\mathbf{b} + \mathbf{Q}'\boldsymbol{\mu} = \boldsymbol{\Omega}\mathbf{b} + (\mathbf{I}_k + \boldsymbol{\Omega}\mathbf{g})\boldsymbol{\mu} = \boldsymbol{\nu}$, implying that

$$f(\mathbf{z}) + \log[p_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})] = -\tfrac{1}{2}\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} + \mathbf{z}'\boldsymbol{\Omega}^{-1}\boldsymbol{\nu} + \alpha'$$

$$= -\tfrac{1}{2}(\mathbf{z} - \boldsymbol{\nu})'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\nu}) + \alpha'', \quad \text{(A8)}$$

where $\alpha''$ is constant as a function of $\mathbf{z}$. The exponent of $e^{f(\mathbf{z})}p_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})$ is thus identical, as a function of $\mathbf{z}$, to that of $p_{\boldsymbol{\nu},\boldsymbol{\Omega}}(\mathbf{z})$. Hence formula 16 holds.

## OTHER ANALYSES THAT CORRESPOND TO LOG-LINEAR AND LOG-QUADRATIC FITNESS FUNCTIONS

In addition to generalised linear models with log link functions, there may be other cases where models of trait-fitness relationships may correspond to log-linear or log-quadratic fitness functions. In this section, we describe how the formulae for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given in equations 17 and 18 are applicable in four kinds of analyses. In each of these situations, it may not be immediately apparent that a form of the fitness function equivalent to equation 9 is implied.

### Models with overdispersion

Key applications of equations to calculate selection gradients from parameters of log-linear fitness models, such as GLM and generalised linear mixed model (GLMM) analysis, are likely to involve fitness measures with more dispersion than is accommodated by standard statistical distributions such as the Poisson distribution. Multiplicative overdispersion of count variables is often handled with a negative binomial distribution, parameterised via its expectation and a dispersion parameter. Thus negative binomial GLM analyses can be used directly to estimate $\mathbf{b}$ and $\mathbf{g}$, and thus $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. For additive overdispersion, the relations between a generalised model of a fitness function and selection gradients are equally direct, but this is not so immediately obvious. Consider a model for expected fitness such as

$$log(E[W]_i) = a + bz_i + \tfrac{1}{2}gz_i^2 + \epsilon_i, \quad \epsilon_i \sim N\big(0, \sigma_\epsilon^2\big),$$

which represents a GLMM analysis with additive overdispersion. Heterogeneity in the $\epsilon_i$ affects the expectation, such that $E[W]_i = e^{a + bz_i + \frac{1}{2}gz_i^2 + \frac{1}{2}\sigma_\epsilon^2}$ (note that the mean of a log normal distribution depends on the dispersion on the log scale such that $\mu_x = e^{\mu_{\log x} + \frac{\sigma_{\log(x)}^2}{2}}$; Aitchison & Brown 1957). In this situation, the selection gradients, $\beta$ and $\gamma$ are again given by equations 17 and 18, even though this expression for log fitness or expected fitness ostensibly differs from that implied by equations 10 a,b. This can be seen by rearranging such that the dependence of expected absolute fitness on the dispersion parameter, $\sigma_\epsilon^2$, is absorbed into the intercept,

$$E[W]_i = e^{\alpha + bz_i + \frac{1}{2}gz_i^2},$$

where

$$\alpha = a + \frac{1}{2}\sigma_\epsilon^2.$$

Since selection gradients do not depend on the intercept $a$ in equations 17 and 18, they do not depend on $\alpha$ in the above expression, and therefore depend neither on the intercept $a$ nor the dispersion parameter $\sigma_\epsilon^2$. Consequently, linear and quadratic terms from log-link GLMMs with additive overdispersion can also be used with equations 17 and 18 to obtain selection gradients for Gaussian traits.

### Multinomial models, as in parentage analysis

In parentage inference, some methods have been proposed wherein the probability that candidate parent $i$ is the parent of a given offspring is modelled according to

$$W(\mathbf{z}_i) \propto e^{f(\mathbf{z}_i)},$$

and where realised parentages of a given offspring array are then modelled according to a multinomial distribution, potentially integrating over uncertainty in paternity assignments based on molecular data (Hadfield et al. 2006, Smouse et al. 1999). When $f(\mathbf{z})$ is a linear function, (Smouse, Meagher & Korbak 1999; T. Meagher, personal communication) interpreted the analysis as being analogous to that of Lande & Arnold (1983), but not necessarily identical. For a linear $f(\mathbf{z})$, this analysis does in fact yield estimates of $\boldsymbol{\beta}$, and for a quadratic function, directional and quadratic selection gradients can be obtained using equations 17 and 18. This can be seen by noting that expected fitness, given phenotype, of candidate fathers for any given offspring array will be, in the log-linear case,

$$W(\mathbf{z}) = ce^{a + \mathbf{b}'\mathbf{z}} \propto e^{\tilde{a} + \mathbf{b}'\mathbf{z}},$$

where $c$ is a constant and $\tilde{a} = a + log(c)$. As $W(\mathbf{z})$ has been expressed in the form of equation 4, it follows from equation 5 that $\boldsymbol{\beta} = \mathbf{b}$. Moreover, if expected parentage is modelled as a log-quadratic function, equations 17 and 18 can be used to recover selection gradients from multinomial parentage models.

## Mark-recapture with constant survival functions

Another case where our formulae may be applicable pertains to inferences of survival rate. Often, data about trait-dependent survival rates may be assessed over discrete intervals. While the experimental unit of time may be an interval (e.g., a day or a year), the biologically-relevant aspect of variation in survival may be longevity, i.e., for how many intervals an individual survives. One such situation arises when per-interval survival rate is assessed via a logistic regression analysis, and trait-dependent survival rates are (or may be assumed to be) constant across intervals. A common case of logistic regression analysis that satisfies this first condition is often implemented in capture-mark-recapture procedures. Suppose that per-interval survival rate, given phenotype, may be assumed to be constant, with death in a particular interval of an individual with phenotype $\mathbf{z}$ occurring at probability $\rho(\mathbf{z})$. If fitness may reasonably be assumed to be reflected by the expected number of intervals survived, then it will be given, as the expectation of a geometric distribution, by

$$W(\mathbf{z}) = \frac{1 - \rho(\mathbf{z})}{\rho(\mathbf{z})}.$$

If trait-dependent per-interval survival probability is denoted $\phi(\mathbf{z})$ ($\phi$ being the standard symbol for survival rate in capture-mark-recapture analyses; Lebreton et al. 1992), then the fitness function in terms of expected number of intervals lived is $W(\mathbf{z}) = \frac{1 - (1 - \phi(\mathbf{z}))}{1 - \phi(\mathbf{z})} = \frac{\phi(\mathbf{z})}{1 - \phi(\mathbf{z})}$. If per-interval survival rate has been modeled as a logistic regression, i.e.,

$$\phi(\mathbf{z}) = \frac{e^{f(\mathbf{z})}}{1 + e^{f(\mathbf{z})}},$$

where $\phi(\mathbf{z})$ denotes the per-interval fitness function, and $f(\mathbf{z})$ is the fitness function on the logistic scale, then the fitness function on the discrete longevity scale is

$$W(\mathbf{z}) = \frac{\frac{e^{f(\mathbf{z})}}{1 + e^{f(\mathbf{z})}}}{1 - \frac{e^{f(\mathbf{z})}}{1 + e^{f(\mathbf{z})}}} = e^{f(\mathbf{z})}.$$

Therefore, if $f(\mathbf{z})$ is a linear function, then its terms are the directional selection gradients on the discrete-longevity scale. If $f(\mathbf{z})$ is a quadratic function, then the corresponding directional and quadratic selection gradients, again if the relevant aspect of fitness is the number of intervals survived, can be obtained using equations 17 and 18. Waller and Svensson (2016) take advantage of these relationships to compare inference of trait-dependent survival in capture-mark-recapture models to classical inference using Lande & Arnold's (1983) least-squares regression analysis

where fitness is assessed as the number of intervals that individuals survive.

It must be stressed that these results do not justify interpretation of logistic regression coefficients of survival probability as selection gradients in a general way. Such coefficients differ from selection gradients for three reasons: (1) they pertain to a linear predictor scale, and natural selection plays out on the data scale, (2) they directly model absolute fitness, not relative fitness, and (3) they pertain to per-interval survival, which may not necessarily be the aspect of survival that best reflects fitness in any given study. It is only when the number of intervals survived is of interest (and phenotype-dependent survival rates can be assumed to be constant across intervals) that these three different aspects of scale cancel out such that the parameters of a logistic regression are selection gradients.

## Survival analysis

Another situation where an important analysis for understanding trait-fitness relationships that has an immediate—but not necessarily immediately apparent—relationship to selection gradients, arises in survival analysis. In a proportional hazards model (Cox 1972), the instantaneous probability of mortality experienced by live individuals, the hazard $\lambda(t)$, as a function of their phenotype, could be modelled as

$$\lambda(t) = \lambda_0 e^{f(z)},$$

where $\lambda_0$ is the baseline hazard, and the factor $e^{f(z)}$ describes individual deviations from this baseline hazard. If the baseline hazard is constant in time, then survival distributions conditional on phenotype are exponential, and have mean $\lambda^{-1}$. So, if fitness may be assumed to be proportional to longevity (as a continuous variable now, not discrete number of intervals as in the relations given above between logistic models of per-interval survival and selection gradients) then

$$W(z) \propto \frac{1}{\lambda_0 e^{f(z)}} = \frac{1}{\lambda_0} e^{-f(z)}.$$

In expressions for selection gradients (equations 1 and 2), $\frac{1}{\lambda_0}$ would be a constant in the integrals in both the numerators and denominators, and therefore cancels in calculations of selection gradients. Therefore, if proportional hazards are modeled with $f(z)$ as a linear or quadratic function, then the expressions for selection gradients (equations 5, 17 and 18) hold, but the coefficients of the trait-dependent hazard function must be multiplied

by -1. The relationship to survival analysis will also approximately hold when fitness is determined by survival to a specific age, and mean survival to that age is low.

## Supporting Information
Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information