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

Michael B. Morrissey[1] and I. B. J. Goudie[2]

February 22, 2016

[1]School of Biology, University of St Andrews

contact
email: michael.morrissey@st-andrews.ac.uk
phone: +44 (0) 1334 463738
fax: +44 (0) 1334 463366
post: Dyers Brae House
School of Biology, University of St Andrews
St Andrews, Fife, UK, KY16 9TH

[2]School of Mathematics and Statistics, University of St Andrews

contact
email: ig@st-andrews.ac.uk
phone: +44 (0) 1334 463705
fax: +44 (0) 1334 463748
post: School of Mathematics and Statistics, University of St Andrews
St Andrews, Fife, UK, KY16 9SS

Keywords: natural selection, selection gradients, fitness, generalised linear model, capture-mark-recapture, survival analysis

Abstract

1. Established methods for inference about selection gradients involve least-squares regression of fitness on phenotype. While these methods are simple and may generally be quite robust, they do not account well for distributions of fitness.

2. Some progress has previously been made in relating inferences about trait-fitness relationships from generalised linear models to selection gradients in the formal quantitative genetic sense. These approaches involve numerical calculation of average derivatives of relative fitness with respect to phenotype.

3. We present analytical results expressing selection gradients as functions of the coefficients of generalised linear models for fitness in terms of traits. The analytical results allow calculation of univariate and multivariate directional, quadratic, and correlational selection gradients from log-linear and log-quadratic models.

4. The results should be quite generally applicable in selection analysis. They apply to any generalised linear model with a log link function. Furthermore, we show how they apply to some situations including inference of selection from (molecular) paternity data, capture-mark-recapture analysis, and survival analysis. Finally, the results may bridge some gaps between typical approaches in empirical and theoretical studies of natural selection.

1 Introduction

The characterisation 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, etc., and ultimately allows meta-analysis (e.g., Kingsolver *et al.* 2001). Selection gradients can characterise both directional selection and aspects of non-linear selection, and so are a very powerful concept in evolutionary quantitative genetics.

Formally, the selection gradient is the vector of partial derivatives of relative fitness with respect to phenotype, averaged over the distribution of phenotype observed in a population. Given an arbitrary function $W(\mathbf{z})$ for expected fitness of a (multivariate) phenotype $\mathbf{z}$, a general expression for the directional selection gradient $\boldsymbol{\beta}$ is

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

where $p(\mathbf{z})$ is the probability density function of phenotype, 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,

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

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).

The primary method for obtaining selection gradient estimates has been a simple and robust approach justified in Lande & Arnold (1983). The method involves least-squares multiple regression of relative fitness, i.e., absolute fitness divided by the mean observed in any comparable group of individuals over a specific period of the life cycle, potentially the entire life cycle, on measures of phenotype. Fitness, or any component of fitness, will typically have highly non-normal residuals in such a regression. Nonetheless, the simple least-squares methods are unbiased (see Geyer & Shaw 2010). However, methods that account for distributions of residuals that arise in regressions involving fitness as a response variable may provide better precision and more reliable statements about uncertainty (i.e., standard errors, p-values, etc.).

Some progress has been made at developing generalised regression model methods for inference of selection gradients. Janzen & Stern (1998) proposed a method for binomial responses (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 standardising 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 characterising 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.

Of the many forms regression analyses of trait-fitness relationships might take, log-linear or log-quadratic models of the relationship between traits and expected absolute fitness may

be particularly useful. In generalised linear models, the log link function is often useful and pragmatic. Fitness is a necessarily non-negative quantity, and expected fitness will typically best be modelled as a strictly positive quantity. This will indeed be the case if expected fitness is an exponential function of the sum of the predictors of the regression model, or, equivalently, a log link is used. Also, a log link function is compatible with generalised linear models with various distributions that could be useful for modelling fitness or fitness components. For example, it can be used with the Poisson distribution (counts, e.g., number of mates or offspring), the negative binomial distribution (for counts that are overdispersed relative to the Poisson distribution, potentially including lifetime production of offspring), and the exponential and geometric distributions (e.g., for continuous and discrete measures of longevity). The purpose of this short paper is to investigate the relationships between log-linear and log-quadratic models of fitness functions, and selection gradients.

2 Log-linear and log-quadratic fitness functions, and selection gradients

Selection gradients turn out to have very simple relationships to the coefficients of log-linear regression models predicting expected fitness from (potentially multivariate) phenotype. Suppose that there are k traits in the analysis and that the absolute fitness function, $W(\mathbf{z})$ takes the form

$$W(\mathbf{z}) = e^{a+b_1 z_1 + b_2 z_2 + \ldots + b_k z_k} \tag{4}$$

where a is a log-scale intercept, and the b_i are log-scale regression coefficients relating the traits (z_i) to expected fitness. The equation for the directional selection gradient (equation 1) can then be simplified. Focusing on the selection gradient for a specific trait, i, in a log-linear model of $W(\mathbf{z})$,

$$\frac{\partial W}{\partial z_i} = b_i W(\mathbf{z})$$

and hence

$$\begin{aligned}\beta_i &= \frac{\int b_i W(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}}{\int W(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}} \\ &= \frac{b_i \int W(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}}{\int W(\mathbf{z}) p(\mathbf{z}) d\mathbf{z}} \\ &= b_i. \end{aligned} \tag{5}$$

This result could be quite useful. In any log-linear model regressing expected absolute fitness, or a component of fitness, on trait values, the linear predictor-scale regression coefficients are the directional selection gradients.

The situation is a little bit more complicated if a log-quadratic model is fitted. If $W(\mathbf{z})$ takes the form

$$W(\mathbf{z}) = e^{a + \sum_i b_i z_i + \sum_i g_i \left(\frac{1}{2} z_i^2\right) + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} g_{ij} (z_i z_j)}, \tag{6}$$

i.e., of a log-scale regression model with linear and quadratic terms, plus first-order interactions, then the b_i coefficients are not necessarily the directional selection gradients, nor are the g_i and g_{ij} coefficients the quadratic and correlational selection gradients, as they would be in a least squares analysis following Lande & Arnold (1983). However, we can use the log-scale quadratic fitness function with the general definitions of selection gradients (equations 1 and 2) to obtain analytical solutions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

The factor of $\frac{1}{2}$ associated with the quadratic terms in equation 6 is a potential source of confusion, analogous to that surrounding a similar factor in Lande & Arnold's (1983) paper (see Stinchcombe *et al.* 2008). In order to obtain the correct values of the g_i coefficients, the covariate values for quadratic terms should be (1) mean-centred, then (2) squared, and then (3) 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 (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 6, and a matrix $\mathbf{g} = \left(g_{ij}\right)$ containing the coefficients of the corresponding quadratic form. We can then write the fitness function more conveniently in matrix form

$$W(\mathbf{z}) = e^{f(\mathbf{z})} \tag{7a}$$

$$f(\mathbf{z}) = a + \mathbf{b}'\mathbf{z} + \frac{1}{2}\mathbf{z}'\mathbf{g}\mathbf{z}. \tag{7b}$$

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 $d_i = E\left[\frac{\partial W(\mathbf{z})}{\partial z_i}\right]$ and the elements of $\mathbf{H}$ are $H_{ij} = E\left[\frac{\partial^2 W(\mathbf{z})}{\partial z_i \partial z_j}\right]$. We can now rewrite the expressions for directional and quadratic selection gradients as

$$\boldsymbol{\beta} = \frac{\mathbf{d}}{E[W(\mathbf{z})]} \tag{8}$$

and

$$\boldsymbol{\gamma} = \frac{\mathbf{H}}{E[W(\mathbf{z})]}. \tag{9}$$

Differentiating equation 7 gives

$$\frac{\partial W(\mathbf{z})}{\partial \mathbf{z}'} = (\mathbf{b} + \mathbf{g}\mathbf{z})\, e^{f(\mathbf{z})}, \tag{10}$$

and

$$\frac{\partial^2 W(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}'} = \left(\mathbf{g} + (\mathbf{b} + \mathbf{g}\mathbf{z})\,(\mathbf{b} + \mathbf{g}\mathbf{z})'\right) e^{f(\mathbf{z})}. \tag{11}$$

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

$$K(\mathbf{z}) = \left(E\left[e^{f(\mathbf{z})}\right]\right)^{-1} e^{f(\mathbf{z})} p_{\mu,\boldsymbol{\Sigma}}(\mathbf{z}) \tag{12}$$

is a probability density function. Define the matrix $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} - \mathbf{g}$ and the vector $\boldsymbol{\nu} = \boldsymbol{\mu} + \boldsymbol{\Omega}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})$. 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 A7,

$$K(\mathbf{z}) \propto p_{\nu,\mathbf{\Omega}}(\mathbf{z}). \tag{13}$$

As K is a probability density function this implies,

$$K(\mathbf{z}) = p_{\nu,\mathbf{\Omega}}(\mathbf{z}). \tag{14}$$

Define $\mathbf{Q}^{-1} = \mathbf{\Omega}^{-1}\mathbf{\Sigma} = \mathbf{I}_k - \mathbf{g}\mathbf{\Sigma}$. Combining equations 8, 10 and 14 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

$$\boldsymbol{\beta} = \mathbf{b} + \mathbf{g}\nu = (\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \mathbf{g}\mathbf{\Omega}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) = (\mathbf{I}_k + \mathbf{g}\mathbf{\Omega})(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) = \mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}), \tag{15}$$

by use of equation A4.

Combining equations 9, 11 and 14 yields $\boldsymbol{\gamma} = E[\mathbf{g} + (\mathbf{b} + \mathbf{g}\mathbf{z})\,(\mathbf{b} + \mathbf{g}\mathbf{z})']$, where the expectation is taken with respect to K. Hence

$$\begin{aligned}
\boldsymbol{\gamma} &= \mathbf{g} + \mathrm{Var}(\mathbf{b} + \mathbf{g}\mathbf{z}) + [\mathrm{E}(\mathbf{b} + \mathbf{g}\mathbf{z})][\mathrm{E}(\mathbf{b} + \mathbf{g}\mathbf{z})]' \\
&= \mathbf{g} + \mathbf{g}\mathbf{\Omega}\mathbf{g}' + \boldsymbol{\beta}\boldsymbol{\beta}' \\
&= \boldsymbol{\beta}\boldsymbol{\beta}' + (\mathbf{I}_k + \mathbf{g}\mathbf{\Omega})\mathbf{g} \\
&= \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{Q}\mathbf{g},
\end{aligned} \tag{16}$$

where we have noted that $\mathbf{g}$ is symmetric and used equation A4.

In univariate analyses, the matrix machinery necessary for implementing the general formulae in equations 15 and 16 can be avoided. If the fitness function is $W(z) = e^{a+bz+g\frac{1}{2}z^2}$ (note, again, that the quadratic coefficient is that for centred, then squared, and then halved values of z[1]), and z has a mean of μ and a variance of σ^2 and 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-standardised,

[1] This can be accomplished easily in `R`. Assume that W and z are variables in memory representing absolute fitness and phenotypic data, and that residuals of W are assumed to follow a Poisson distribution. The regression could be implemented by `glm(W~z+I(0.5*(z-mean(z))^2),family=poisson(link="log"))`.

variance-standardised, and unstandardised selection gradients, when appropriate values of μ and σ^2 are used, and applied to log-quadratic models of $W(z)$ where the phenotypic records have been correspondingly standardised. For the common case where the trait is mean-centred and (unit) variance standardised, the expressions simplify further to $\beta = \frac{b}{1-g}$ and $\gamma = \frac{b^2+g(1-g)}{(1-g)^2}$.

The equivalence of the regression coefficients of a log-linear fitness model with directional selection gradients (equation 5) of course requires that the regression model provides a reasonable description of the relationship between a trait and expected fitness and makes reasonable assumptions about fitness residuals. Otherwise, the relationship is relatively unburdened by assumptions. For example, it does not require any specific distribution of phenotype. The use of selection gradients obtained from log-linear regressions to predict evolution using the Lande equation (equation 3) does assume that breeding values are multivariate normal (see Morrissey 2014 for a discussion of selection gradients and associated assumptions about multivariate normality of phenotype and breeding values). The expressions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ given a log-quadratic fitness model (equations 15 and 16) do assume multivariate normality of phenotype. Equations 15 and 16 further require that $\boldsymbol{\Omega}$ is positive definite. In univariate analyses, this condition reduces to $g < \frac{1}{\sigma^2}$, implying that the fitness function should not curve upwards too sharply within the range of observed phenotype.

A very convenient feature of the expressions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in equations 5, 15 and 16 is that the model (log) intercept does not influence the selection gradients. This means that the range of modelling techniques that yield selection gradients can be even further expanded. For example, adding fixed and random effects to Lande & Arnold's (1983) least squares analysis will generally result in estimated regression coefficients that are not interpretable as selection gradients. For example, it might be desirable to estimate a single selection gradient across two sexes, if data are limited and sex-differences in selection are not anticipated. In such an analysis, it might seem sensible to include an effect of sex, to account for differences in mean fitness between the sexes. However, such an analysis would not yield correct selection gradients, because the theory underlying the least squares-based regression analysis of selection requires that mean relative fitness is one, and this would not be the case when different strata within an analysis have different intercepts. On the other hand, adding such an effect to a log-scale model of absolute

fitness, and then deriving selection gradients using equations 5, 15 and 16 will yield correct selection gradients. Other effects, such as random effects to account for individual heterogeneity in expected fitness, beyond that explained by the traits (or correlated, unmeasured traits), will be usable as well, while still retaining the ability to obtain correct selection gradients.

3 Statistical uncertainty

The expressions for selection gradients, given the parameters of a log-quadratic fitness function (equations 15 and 16) give the selection gradients conditional on the estimated values of **b** and **g**. However, **b** and **g** will not typically be known quantities in empirical studies of natural selection, but rather will be estimates with error. Because equations 15 and 16 are non-linear functions of one or more regression coefficients, unconditional estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ would have to be obtained by integrating the expressions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ over the sampling distributions of the estimated values of **b** and **g**. Such details are not normally considered in calculations of derived parameters (e.g., heritabilities) in evolutionary studies. Such integration could be achieved using approximations, bootstrapping, or MCMC methods. Alternatively, application of equations 15 and 16 directly to estimated values of **b** and **g** may be sufficient in practice. Similarly, while standard errors of the parameters **b** and **g** are not directly interpretable as standard errors of corresponding values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, approximations, bootstrapping, and MCMC methods may all potentially be useful in practice. In particular, approximation of standard errors by a first-order Taylor approximation (the "delta method"; Lynch & Walsh 1998) may generally be pragmatic. Formulae for approximate standard errors by this method are given in the appendix. For univariate analysis, with phenotype standardised to $\mu = 0$ and $\sigma^2 = 1$, the approximate standard errors of β and γ are given by

$$SE[\beta] \approx \sqrt{\frac{\Sigma[b]}{(1-g)^2} + \frac{b^2\Sigma[g]}{(1-g)^4} + \frac{2b\Sigma[b,g]}{(1-g)^3}}, \tag{17}$$

and

$$SE[\gamma] \approx \sqrt{\frac{4b^2\Sigma[b]}{(1-g)^4} + \frac{(1+2b^2-g)^2\,\Sigma[g]}{(1-g)^6} + \frac{4b(1+2b^2-g)\Sigma[b,g]}{(1-g)^5}}. \tag{18}$$

Where $\Sigma[b]$ and $\Sigma[g]$ represent the sampling variances of the estimated b and g terms. These are the squares of their standard errors. $\Sigma[b,g]$ is the sampling covariance of the b and g terms. This is not always reported, but can usually be obtained. For example, in `R`, it can be extracted from a fitted glm object using the function `vcov()`.

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 their standard errors. We simulated univariate directional selection, with values of b between -0.5 and 0.5, and with $g = -0.5, 0$ and 0.2. Because β and γ are non-linear 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 γ. 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 non-linear 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.

Firstly we analysed each simulated dataset using the OLS regression described by Lande & Arnold (1983), i.e., $w_i = \mu + \beta z_i + \gamma\left(\frac{1}{2}z_i^2\right) + e_i$, 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. Secondly we fitted a Poisson glm with a linear and quadratic terms, using the `R` function `glm()`. We then calculated conditional selection gradient estimates using equations 15 and 16. 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 β and γ 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 β and γ in our simulations. Similarly, first-order approximations of standard errors of the glm-derived estimates of β and γ closely reflected the simulated standard deviations of the estimators (figure 1). All methods for obtaining standard errors performed well for estimates of β 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 standard OLS standard errors. Mean absolute error of the glm estimators was always smaller than that of the OLS estimators of β and γ. This is unsurprising, as the simulation scheme corresponded closely to the glm model. These results demonstrate the usefulness of the conditional values of β and γ 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.

4 Other analyses that correspond to log-linear 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 paternity inference, some methods have been proposed wherein the probability that candidate father i is the father of a given offspring is modelled according to

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

and where realised paternities 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 $f(z)$ is a linear function, Smouse, Meagher & Korbak (1999; T. Meagher, personal communication) interpreted the analysis as analogous to Lande and Arnold's 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 15 and 16. 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,

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

where c is a constant. In application of the expressions yielding equation 5, c appears in both the numerator and the denominator, yielding $\boldsymbol{\beta} = \mathbf{b}$.

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, and that fitness is defined to be the expected survival time. Then fitness will be given by the mean of a geometric distribution where death in a particular interval of an individual with phenotype $\mathbf{z}$ occurs with probability $\rho(\mathbf{z})$,

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

If trait-dependent per-interval survival probability is denoted $\phi(\mathbf{z})$ (ϕ 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

279 survival rate has been modelled as a logistic regression, i.e.,

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

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

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

282 Therefore, if $f(\mathbf{z})$ is a linear function, then its terms are the directional selection gradients on
283 the discrete-longevity scale. If $f(\mathbf{z})$ is a quadratic function, then the corresponding directional
284 and quadratic selection gradients, again if the relevant aspect of fitness is the number of
285 intervals survived, can be obtained using equations 15 and 16. Waller and Svensson (2016; this
286 issue) takes advantage of these relationships to compare inference of trait-dependent survival
287 in capture-mark-recpature models to classical inference using Lande & Arnold's (1983) least-
288 squares regression analysis where fitness is assessed as the number of intervals that individuals
289 survive.

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

299 Finally, another situation where an important analysis for understanding trait-fitness
300 relationships that has an immediate – but not necessarily immediately apparent – relationship
301 to selection gradients, arises in survival analysis. In a proportional hazards model (Cox, 1972),
302 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 λ_0 is the baseline hazard, and the $e^{f(z)}$ part of the function 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 λ^{-1}. So, if fitness is taken to be expected 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

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

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 modelled with $f(z)$ as a linear or quadratic function, then the expressions for selection gradients (equations 5, 15 and 16) hold, but the coefficients of the trait-dependent hazard function must be multiplied by -1.

5 Conclusion

We have provided analytical expressions for selection gradients, given the parameters of log-linear and log-quadratic functions describing expected fitness. These functions can be applied in conjunction with a range of generalised linear model approaches, specific situations in capture-mark-recapture analysis, and relate to fitness functions used in theoretical studies. The general relationship of selection gradients to the coefficients of log-linear and log-quadratic models, in particular, various generalised 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 (e.g., reasonable standard errors) than the main methods used to date. 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. In theoretical studies, Gaussian fitness functions are often used. These are simply log-quadratic functions that are parameterised in terms of a location parameter (phenotype of maximum fitness), and a width parameter. A relationship between the parameters of a Gaussian fitness function and directional selection gradients (Lande 1979; the expression we give for $\boldsymbol{\beta}$ is an alternative formulation) is already widely used in the theoretical literature. For any given distribution of phenotype, these parameters correspond directly to linear and quadratic (log-scale) regression parameters, and so can be directly related to selection gradients in empirical studies.

Acknowledgements

We thank Andy Gardner, Graeme Ruxton, and Kerry Johnson for discussions, comments, and advice. Peter Jupp provided particular insights that improved this paper. MBM is supported by a Royal Society (London) University Research Fellowship.

References

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.

Geyer, C.J. & Shaw, R.G. (2010) Aster models and the Lande-Arnold beta. Technical report, University of Minnesota.

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

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*. Sinauer, Sunderland, MA.

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

Morrissey, M.B. (2014) In search of the best methods for multivariate selection analysis. *Methods in Ecology and Evolution*, **5**, 1095–1109.

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

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: Dougle 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*.

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

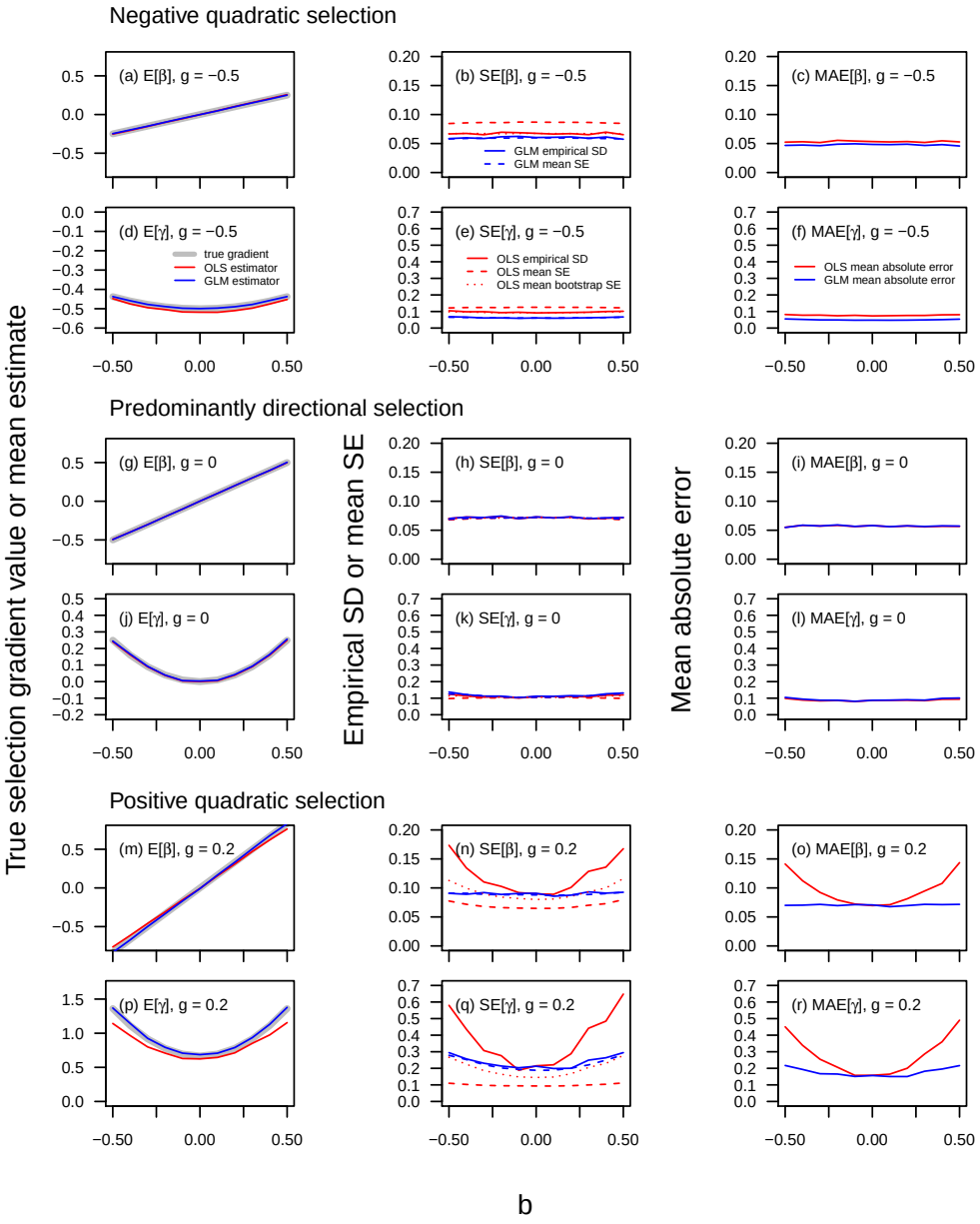

Figure 1: Simulation results for the performance of Lande & Arnold's (1983) least squares-based (OLS) estimators (red lines), and log-quadratic (GLM) estimators (blue lines), of directional and quadratic selection gradients. The first column shows bias in estimates of β and γ, where departure from the 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 absolute errors of the OLS and GLM estimators.

Appendix

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, $\tilde{\boldsymbol{\gamma}} = \left[\gamma_{1,1}, \gamma_{2,1}, \gamma_{3,1}, \gamma_{2,2}, \gamma_{3,2}, \gamma_{3,3}\right]'$. 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 6 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 $\boldsymbol{\beta} = \mathbf{Q}(\mathbf{b} + \mathbf{g}\mu)$ (equation 15),

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{b}} = \mathbf{Q}. \tag{A1}$$

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

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

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 $\mathbf{J}$.

$$\begin{aligned}
\boldsymbol{\beta} = \boldsymbol{\Psi}^{-1}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) \qquad \Rightarrow \frac{\partial \boldsymbol{\beta}}{\partial \tilde{g}_m} &= \frac{\partial \boldsymbol{\beta}}{\partial g_{ij}} = -\boldsymbol{\Psi}^{-1}\left[\frac{\partial \boldsymbol{\Psi}}{\partial g_{ij}}\right]\boldsymbol{\Psi}^{-1}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \boldsymbol{\Psi}^{-1}\left[\frac{\partial (\mathbf{b} + \mathbf{g}\boldsymbol{\mu})}{\partial g_{ij}}\right] \\
&= -\mathbf{Q}\left[\frac{\partial\, \mathbf{I}_k - \mathbf{g}\boldsymbol{\Sigma}}{\partial g_{ij}}\right]\mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu}) + \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right]\boldsymbol{\mu} \\
&= \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right][\boldsymbol{\Sigma}\mathbf{Q}(\mathbf{b} + \mathbf{g}\boldsymbol{\mu})] + \mathbf{Q}\left[\frac{\partial \mathbf{g}}{\partial g_{ij}}\right]\boldsymbol{\mu} \\
&= \mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu}) = \mathbf{Q}\mathbf{C}_m(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu}) \\
\Rightarrow \frac{\partial \boldsymbol{\beta}}{\partial \tilde{\mathbf{g}}} &= \sum_{m=1}^{s} \frac{\partial \boldsymbol{\beta}}{\partial \tilde{g}_m}\mathbf{e}'_m = \mathbf{Q}\sum_{m=1}^{s}\mathbf{C}_m(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})\mathbf{e}'_m
\end{aligned} \tag{A2}$$

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 $\mathbf{J}$.

$$\begin{aligned}
\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{Q}\mathbf{g} \qquad \Rightarrow \frac{\partial \boldsymbol{\gamma}}{\partial b_u} &= \boldsymbol{\beta}\left(\frac{\partial \boldsymbol{\beta}}{\partial b_u}\right)' + \left(\frac{\partial \boldsymbol{\beta}}{\partial b_u}\right)\boldsymbol{\beta}' = \boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}' \\
\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial b_u} &= \mathbf{v}\left(\boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}'\right) \\
\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \mathbf{b}} &= \sum_{u=1}^{k}\mathbf{v}\left(\boldsymbol{\beta}\mathbf{Q}'_{[u]} + \mathbf{Q}_{[u]}\boldsymbol{\beta}'\right)\mathbf{e}'_u
\end{aligned} \tag{A3}$$

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

$$\mathbf{I}_k + \mathbf{g}\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega} = \mathbf{Q} \tag{A4}$$

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

$$\mathbf{Q}' = \mathbf{I}_k + (\mathbf{g}\boldsymbol{\Omega})' = \mathbf{I}_k + \boldsymbol{\Omega}\mathbf{g}. \tag{A5}$$

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

$$\begin{aligned}
\frac{\partial \gamma}{\partial g_{ij}} &= \left[\frac{\partial \boldsymbol{\beta}}{\partial g_{ij}}\right]\boldsymbol{\beta}' + \boldsymbol{\beta}\left[\frac{\partial \boldsymbol{\beta}}{\partial g_{ij}}\right]' + \mathbf{Q}\mathbf{C}^{(ij)} + \mathbf{Q}\mathbf{C}^{(ij)}\boldsymbol{\Sigma}\mathbf{Q}\mathbf{g} \\
&= \left[\mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})\right]\boldsymbol{\beta}' + \boldsymbol{\beta}\left[\mathbf{Q}\mathbf{C}^{(ij)}(\boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\mu})\right]' + \mathbf{Q}\mathbf{C}^{(ij)} + \mathbf{Q}\mathbf{C}^{(ij)}\boldsymbol{\Omega}\mathbf{g} \\
\Rightarrow \frac{\partial \tilde{\gamma}}{\partial g_{ij}} &= \mathbf{v}\left[\mathbf{M}^{(m)} + (\mathbf{M}^{(m)})' + \mathbf{Q}\mathbf{C}_m(\mathbf{I}_k + \boldsymbol{\Omega}\mathbf{g})\right] \\
\Rightarrow \frac{\partial \tilde{\boldsymbol{\gamma}}}{\partial \tilde{\mathbf{g}}} &= \sum_{m=1}^{s}\mathbf{v}\left[\mathbf{M}^{(m)} + (\mathbf{M}^{(m)})' + \mathbf{Q}\mathbf{C}_m\mathbf{Q}'\right]\mathbf{e}'_m,
\end{aligned} \tag{A6}$$

by use of equation A5.

Finally note that equations A4 and A5 are also relevant to the derivation of formula 13. 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 α does not depend on $\mathbf{z}$. Thus, if $\alpha' = \alpha + a$, it follows that, as a function of $\mathbf{z}$,

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

Now, by A4 and A5, 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

$$f(\mathbf{z}) + \log[p_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{z})] = -\tfrac{1}{2}\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} + \mathbf{z}'\boldsymbol{\Omega}^{-1}\boldsymbol{\nu} + \alpha' = -\tfrac{1}{2}(\mathbf{z} - \boldsymbol{\nu})'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\nu}) + \alpha'', \tag{A7}$$

where α'' is constant as a function of $\mathbf{z}$. The exponent of $e^{f(\mathbf{z})}p_{\mu,\boldsymbol{\Sigma}}(\mathbf{z})$ is thus identical, as a function of $\mathbf{z}$, to that of $p_{\nu,\boldsymbol{\Omega}}(\mathbf{z})$. Hence formula 13 holds.