Syst. Biol. 61(3):413–425, 2012

DOI:10.1093/sysbio/syr122
Advance Access publication on January 2, 2012

Interpreting the Evolutionary Regression: The Interplay Between Observational and Biological Errors in Phylogenetic Comparative Studies

THOMAS F. HANSEN[1,*] AND KRZYSZTOF BARTOSZEK[2]

[1]*Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, PB 1066, Blindern, N-0316 Oslo, Norway; and*
[2]*Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, SE-412 96 Gothenburg, Sweden;*
[*]*Correspondence to be sent to: Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, PB 1066, Blindern, N-0316 Oslo, Norway; E-mail: Thomas.Hansen@bio.uio.no.*

Received 28 March 2011; reviews returned 21 July 2011; accepted 12 October 2011
Associate Editor: Todd E. Oakley

Abstract.—Regressions of biological variables across species are rarely perfect. Usually, there are residual deviations from the estimated model relationship, and such deviations commonly show a pattern of phylogenetic correlations indicating that they have biological causes. We discuss the origins and effects of phylogenetically correlated biological variation in regression studies. In particular, we discuss the interplay of biological deviations with deviations due to observational or measurement errors, which are also important in comparative studies based on estimated species means. We show how bias in estimated evolutionary regressions can arise from several sources, including phylogenetic inertia and either observational or biological error in the predictor variables. We show how all these biases can be estimated and corrected for in the presence of phylogenetic correlations. We present general formulas for incorporating measurement error in linear models with correlated data. We also show how alternative regression models, such as major axis and reduced major axis regression, which are often recommended when there is error in predictor variables, are strongly biased when there is biological variation in any part of the model. We argue that such methods should never be used to estimate evolutionary or allometric regression slopes. [Adaptation; allometry; major-axis regression; measurement error; phylogenetic comparative method; phylogenetic inertia; reduced major-axis regression; structural equation.]

Evolutionary regression across species is one of the major statistical procedures used to study the evolutionary relationship between biological variables and to test hypotheses about adaptation to environmental variables (Harvey and Pagel 1991). Over the last decades, sophisticated statistical models have been developed to deal with the problem of phylogenetic correlations between related species, the incorporation of past history, and issues having to do with observation error and scaling. Most of these developments have been focused on solving statistical problems, and the biological interpretations and implications of the methods have often been ignored. Assumptions made to solve statistical problems are often incompatible with the biological processes that motivated the approaches. For example, the almost universal assumption that the residuals of a phylogenetic regression evolve as Brownian motion is made for analytical convenience and is inconsistent with adaptive evolution (Hansen and Orzack 2005), and many approaches to deal with the bias that results from observation error in evolutionary and allometric regression are based on statistical models that effectively assume that there is no biological error in the fitted model. In this paper, we will build on the distinction between biological and observational error introduced by Riska (1991) and study their joint effects in evolutionary regression models. We will argue that both biological and statistical considerations are necessary to obtain meaningful evolutionary and allometric regression models.

Observation or measurement error is a serious concern for most comparative studies (e.g., Pagel and Harvey 1988; Kelly and Price 2004; Ives et al. 2007). Often, species data are means of variables that are computed from small samples of individuals. Apart from errors stemming from nonrandom sampling of individuals from a species, such means have an estimation error, usually quantified by the standard error of the estimate, which will act as a "measurement error" in the comparative analysis. When such observation error can be quantified, it should be incorporated into the analysis to improve precision and to correct possible bias in the estimated regression slopes. It is, however, instrumental to separate observation error from biological "error," which may be defined as the true "biological" deviance from the assumed model relationship. Such biological deviations will almost always be present simply because biology is a complex affair. We may postulate simple single-factor hypotheses about trait Y being adapted to environment X and take a nonzero regression of Y on X as evidence for this hypothesis, but no real biologist would take this to mean that X is the only biological source of variation in Y. Instead, the biologist will expect considerable variation in Y being due to other, usually unknown, biological sources. Indeed, it is an implicit premise of phylogenetic comparative methods that most or all the residual variation is of biological origin because if deviance from the model was due to observational error, we would not expect to see phylogenetically correlated residuals and thus not need to use phylogenetic comparative methods at all.

We will start by presenting a biological model for the evolutionary regression, which can accommodate either adaptive evolution or allometric constraints, and then combine this with a model of observation error. We will quantify the biases that arise from both biological and observational deviations from the regression model and show how these can be corrected in the presence of phylogenetic correlations. Finally, we will show that biological variation makes major axis and reduced major axis regression strongly biased and unsuitable

as estimators of the evolutionary regression parameter as defined in our model. We also present general formulas for correction of bias due to measurement error in linear models with correlated data.

Biological Models for the Evolutionary Regression

Consider a biological trait Y that differs among a set of species. The causes of such differences are usually manifold. Natural selection will operate on Y, but the exact optimum value will depend on numerous ecological factors that all vary among species in different patterns. The selection on Y will also be influenced by the states of other functionally related characters due to correlated selection, and there will be many sources of indirect selection acting on Y due to correlations with other traits under selection. Species differences in Y could also arise due to genetic drift and environmental plasticity. Assuming we knew the exact effects of all these factors in a species, we could make an exact prediction of Y for this species, as $Y = f(X_1, \ldots, X_m)$, for some function f, where the X_i are the exact states of the relevant factors in this species, but in reality, we will have incomplete information about the states of the factors. Usually, an evolutionary regression is done to test the influence of one or a few focal variables for which we know the state in each species. If X_1 is a such focal variable, the evolutionary regression takes the form

$$Y = \beta_0 + \beta_1 X_1 + r(X_1, ..., X_m), \quad (1)$$

where $r(X_1, \ldots, X_m) = f(X_1, \ldots, X_m) - (\beta_0 + \beta_1 X_1)$ are the biological residuals of the model. Obviously, these residuals will be different in different species due to different states of the X variables. It seems reasonable to assume that there will be a large number of such variables with mostly small differences between species, and the central limit theorem then suggests that the residuals may approach a normal distribution across species, as assumed by the standard linear regression model. There is a complication, however, in that related species may tend to have similar states of some of their X variables, thereby giving rise to phylogenetic correlations in the residuals. Fundamentally, this is the justification for use of phylogenetic corrections in cross-species regression models (e.g., Grafen 1989; Ridley 1989; Nee et al. 1996; Hansen 1997; Price 1997).

To use phylogenetic comparative methods, it is necessary to model the pattern of phylogenetic correlations in the residuals. This requires a phylogeny with branch lengths and a process model of evolutionary change (Hansen and Martins 1996). The standard choice for this model has been Brownian motion, which may seem like a reasonable choice if there are many X-factors that can change at random in any small part of the phylogeny (e.g., Grafen 1989). The Brownian-motion model predicts that the covariance between any two species' residuals will be proportional to their shared phylogenetic branch length and this makes it easy to correct the regression by use of generalized least squares (GLS) or maximum-likelihood techniques (Felsenstein 1985; Grafen 1989; Lynch 1991; Martins and Hansen 1997).

The Brownian-motion model can describe correlated evolution between traits (Felsenstein 1985), but this is only one way an evolutionary regression can arise, and it is not consistent with adaptation of trait to an independent variable (Hansen and Orzack 2005). In fact, residuals representing biological deviations from model predictions may have complex correlation structures that cannot be separated from the prediction model. We illustrate this by discussing two specific models for the evolutionary regression, one based on adaptation and the other on allometric constraints.

The Evolutionary Regression as an Adaptation

An evolutionary process of adaptation implies a tendency for Y to evolve towards an optimal state predicted by X. Residual deviations then represent maladaptation and cannot follow Brownian motion, which does not allow any systematic decrease over time. A simple stochastic process model that does allow a systematic tendency for Y to evolve towards an optimum is the Ornstein–Uhlenbeck process, which models the change in Y in a infinitesimal time interval dt, as

$$dY = -\alpha(Y - \theta)dt + \sigma dW, \quad (2)$$

where dW is white noise representing change in unobserved residual variables, σ is the standard deviation of these changes, and $\alpha > 0$ is the "rate of adaptation" towards the optimum θ, which we model as a linear "optimal" regression, $\theta = b_0 + b_1 X$. The exact regression model predicted by this will depend on how X evolves. Hansen (1997) derived a weighted regression based on a priori information on how X was distributed on a phylogeny such that states of X further down in the phylogeny are discounted with a factor that depends on α (see also Butler and King 2004). In this model, the residual covariances will decrease exponentially with phylogenetic distance at a rate proportional to α. Hansen et al. (2008) derived an evolutionary regression from the assumption that X changes like Brownian motion. In this case, the residual covariances follow a complex function of the phylogeny that also depends on the regression parameter b_1 (see Hansen et al. 2008 for exact equation). Importantly, the evolutionary regressions predicted by these models are more shallow (closer to zero) than the optimal regression slope, b_1. More precisely, the predicted evolutionary regression is

$$Y = \beta_0 + \rho b_1 X, \quad (3)$$

where $\rho = 1 - (1 - e^{-\alpha t})/\alpha t$ is a "phylogenetic correction factor" that depends on α and on the distance, t, from the start of the adaptive radiation to the observed tip. This ρ approaches unity when the rate of adaptation, α, is very high, but if the rate of adaptation is slow, the evolutionary regression slope, $\beta_1 = \rho b_1$, becomes

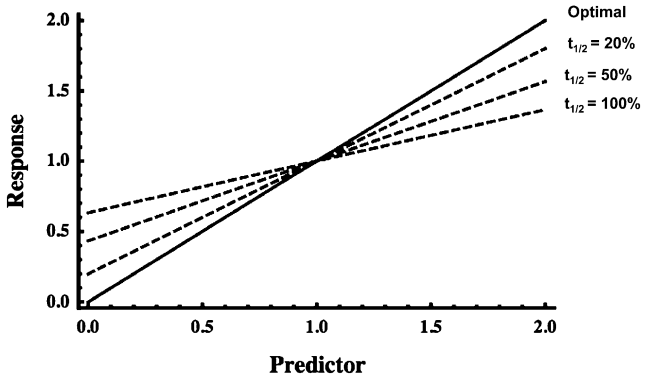

FIGURE 1. Effects of phylogenetic inertia on the expected regression slope. The solid line is an assumed optimal regression in which species are uninfluenced by ancestry. The dashed lines show the expected evolutionary regression slope with increasing levels of phylogenetic inertia measured as a phylogenetic half-life ($t_{1/2} = \text{Log}[2]/\alpha$), which increases from $t_{1/2} = 20\%$ through $t_{1/2} = 50\%$ to $t_{1/2} = 100\%$ of the distance to the root of the phylogeny as we go from the steeper to the most shallow dashed line. The corresponding values of the phylogenetic correction factors are $\rho = 0.80$, $\rho = 0.57$, and $\rho = 0.37$, respectively. A value $t_{1/2} = 100\%$ means that a species that has evolved in a constant environment from the root of the tree is expected to be midway between its ancestral value and its optimum in the constant environment.

shallower than the optimal slope, b_1, as illustrated in Figure 1. The intercept β_0 is also a complex function of the model parameters and not generally equal to the optimal intercept, b_0.

In summary, the evolutionary regressions predicted from simple models of systematic evolutionary adaptation of a dependent variable (= trait) to an independent predictor variable have two features not found in standard phylogenetic regression models. First, residual covariances are not proportional to phylogenetic branch lengths but follow more complex functions dominated by an exponential decrease that depends on both the rate of adaptation and the slope of the regression. Second, the evolutionary regression depends on the rate of adaptation and is more shallow than the optimal relation between the variables. These two features have a common cause in maladaptation of the species. If species cannot adapt instantaneously, then they will deviate from the optimal relationship in the direction of their ancestors, thus producing a more shallow evolutionary regression, and related species will deviate in a similar way, thus causing phylogenetically correlated residuals. In this situation, the residual covariance structure should not be modeled independently of the regression, as done in essentially all phylogenetic regression models used to date.

The Evolutionary Regression as an (Allometric) Constraint

Alternatively, a cross-species regression may reflect a constraint on the independent evolution of the two variables. This is perhaps most obvious in the case of allometric relationships. When size-related traits are plotted against each other on a log-log scale, they often follow a linear relation both within and among species. A cross-species (evolutionary) allometry can arise by at least two distinct mechanisms. One hypothesis is that it reflects adaptation to a functional scaling relation between the two traits. In this case, the observed evolutionary allometries may be shallower than the optimal relation between the traits, as described by the adaptation model above. Another hypothesis is that the evolutionary allometry results from the constraints of an underlying within-species (static) allometry. Static allometries can result from common growth regulation (Huxley 1932; Savageau 1979), and if this regulation is fixed, the traits become constrained to change in concert. This differs from the adaptation model in that no evolutionary deviation from the regression line is possible, and there is no meaningful distinction between an optimal and an evolutionary regression.

The question that arises is how to interpret residual deviations under the constraint model. Observation error will be one source of deviation and may bias the slope as described later. Biological error may also arise if the allometric relation is not perfect. Consider a trait Y that consists of two components, one that depends in a fixed manner on a trait X and one that depends on a set of other unknown variables as above. As an example, if Y is the volume of the organism and X its length, then it is reasonable to expect that any change in X would be accompanied by an automatic instantaneous change in Y, but Y could also change due to evolutionary changes in shape of the organism, and this could cause deviations from the (log-log) regression of Y on X. This leaves open the evolutionary dynamics of the residuals. If shape changes are evolutionarily independent from the organism's size, then these deviations could behave like Brownian motion, and their covariances could be proportional to shared branch lengths. Perhaps more realistically, shape could tend to evolve around an optimum and make the the residuals better represented by an Ornstein–Uhlenbeck process, but this would not be coupled with a biased regression slope as in the adaptation model.

THE STATISTICAL MODEL

The standard setup for the evolutionary regression model is

$$\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \mathbf{r}, \mathbf{r} \sim \text{N}(\mathbf{0}, \mathbf{V}), \tag{4}$$

where $\mathbf{y}$ is a vector of dependent variables, $\mathbf{D}$ is a design matrix containing the predictor variables, $\boldsymbol{\beta}$ is a vector of regression parameters to be estimated (including an intercept), and $\mathbf{r}$ are residuals assumed to be normally distributed with mean zero and variance matrix, $\mathbf{V}$. If the residuals are phylogenetically correlated, these correlations are modeled as off-diagonal elements in the $\mathbf{V}$ matrix, and an unbiased

minimum variance estimator of $\boldsymbol{\beta}$ is then the GLS estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\ \mathbf{D}^T\mathbf{V}^{-1}\mathbf{y}. \quad (5)$$

Note that both the residual variance matrix and the design matrix may depend on unknown parameters, as α in the adaptation model. This can be solved either by estimating all parameters jointly by maximum likelihood or by estimated GLS where equation (5) is used to estimate $\boldsymbol{\beta}$ based on an initial assumption about the other parameters, which are then estimated by maximum likelihood conditionally on $\boldsymbol{\beta}$, whereupon the scheme is iterated until convergence (e.g., Martins and Hansen 1997).

Observation Error

In an evolutionary regression analysis, both dependent and predictor variables are often vectors of trait means or other sample statistics from the individual species. These statistics are necessarily subject to sampling error, which will act as observation (measurement) error in the regression analysis. Such observation errors will reduce the precision of parameter estimates, make confidence intervals too narrow, and bias regression slopes when they occur in predictor variables (Fuller 1987; Buonaccorsi 2010). Due to the small sample sizes for individual species in many comparative analyses, these problems can be substantial. Many comparative studies also have very different sample sizes for different species and can be improved by weighing the species with their reliability.

We may illustrate these effects with a simple model with one predictor variable (as in Riska 1991; Kelly and Price 2004). Let y_o and y_t be the observed and true values of the response variable, and let x_o and x_t be the observed and true values of the predictor variable. Then,

$$y_o = y_t + e, \quad (6a)$$

$$x_o = x_t + u, \quad (6b)$$

where e and u are random observation errors. As in the "classical model" of measurement error (Fuller 1987; Buonaccorsi 2010), we assume that e and u are independent of y_t and x_t, and note that this makes them correlated with y_o and x_o. By assumption, the relationship between the true values of the two variables is

$$y_t = \beta_0 + \beta_1 x_t + r_t, \quad (7)$$

where β_0 and β_1 are parameters modeling the predicted biological relationship between y_t and x_t, and r_t is the true (biological) species deviance from the predicted relationship. From the above, the relationship between the observed variables will be

$$y_o = y_t + e = \beta_0 + \beta_1 x_t + r_t + e = \beta_0 + \beta_1 x_o - \beta_1 u + r_t + e. \quad (8)$$

Thus, we see that the overall residual of the regression of observed response variables on observed predictor variables has three components:

$$r = r_t + e - \beta_1 u. \quad (9)$$

Consequently, the residual variance matrix of this regression model also has three components:

$$\mathbf{V} = \mathbf{V}_t + \mathbf{V}_e + \beta_1^2\mathbf{V}_{u|x}, \quad (10)$$

which we write as matrices (bold face) to allow for covariances between the residuals. The matrix $\mathbf{V}_t$ is the biological variance around the predicted relationship between the true variables, and this is the component that we expect to have a phylogenetic structure (i.e., off-diagonal covariances derived from the assumed model of evolution). The observation variance in the response variable, $\mathbf{V}_e$, can be added to this. The effects of observation variance in the predictor are more complicated. First, since the observation error in the predictor is correlated with the observed predictor, we should not use the raw observation variance, $\mathbf{V}_u$, but rather $\mathbf{V}_{u|x}$, the observed variance conditional on the x_o (computation of this is given below). Second, this variance matrix needs to be multiplied with the square of the regression slope, and this suggests an iterated estimation procedure, which can be incorporated into the estimated GLS procedure discussed above.

Including the observation variance by basing the GLS on equation (10) improves precision of estimates and gives more correct confidence intervals, but it does not correct the bias in the regression slope that results from error in the predictor variables. In the appendix, we show that conditioning on the observed predictor variables, and assuming a normal distribution of the observation error in the predictors, the mean and variance of the estimated regression slope $\hat{\beta}_1$ are

$$E[\hat{\beta}_1\ |\mathbf{x}_o] \approx \left(1 - \frac{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{V}_u\mathbf{V}_x^{-1}\mathbf{x}_o}{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o}\right)\beta_1, \quad (11a)$$

$$\mathrm{Var}[\hat{\beta}_1|\mathbf{x}_o] \approx \frac{1}{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o}, \quad (11b)$$

where $\mathbf{x}_o$ is the vector of observed predictor variables, which we assume are centered on their means, $\mathbf{V}_x$ is their variance matrix, including the observation variance, $\mathbf{V}_u$, and $\mathbf{V} = \mathbf{V}_t + \mathbf{V}_e + \hat{\beta}_1^2\mathbf{V}_{u|x}$ is the total residual variance with $\mathbf{V}_{u|x} = \mathbf{V}_u - \mathbf{V}_u\mathbf{V}_x^{-1}\mathbf{V}_u$. Thus, the larger the observation variance in the predictor variables, the more the estimated slope is expected to diverge from the true slope (Fig. 2). Usually, the estimated slope is too shallow and tends to get increasingly shallow as observation variance increases relative to the total variance in the predictor. The bias disappears when $\mathbf{V}_u$ approaches zero, thus confirming that the GLS estimator is unbiased with observation error only in the dependent variable. If both the observation errors and the predictor variables are independent and homoscedastic, that

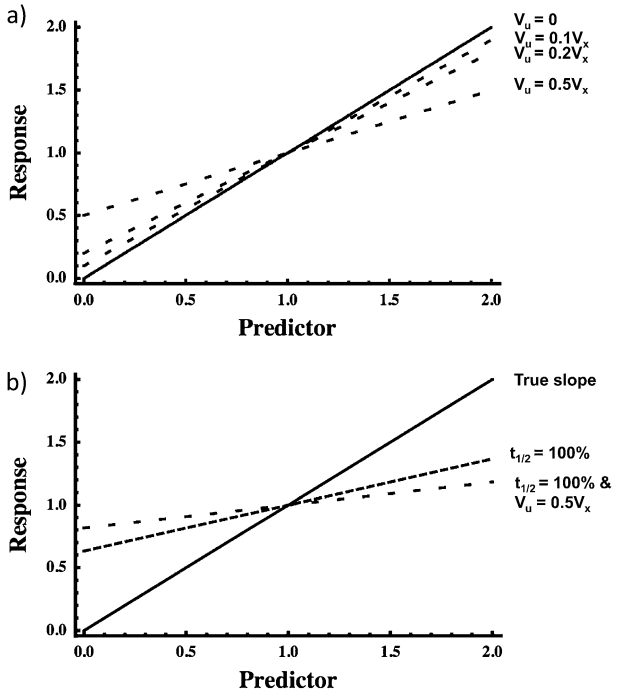

FIGURE 2. Effects of observation error in the predictor variable on the expected regression slope. The solid line is the true regression. In (a), the dashed lines show the effects of letting 10%, 20%, and 50% of the variance in the predictor variable be due to observation error. In (b), we show the effects of combining phylogenetic inertia and observation error in the predictor. The steeper dashed line is the expected evolutionary regression with $t_{1/2} = 100\%$, as in Figure 1, and the shallower dashed line shows the effects of making 50% of the predictor variance due to error. The combined effect is a slope only 18.5% of the optimal and "error-free" regression slope. For ease of illustration, we have assumed that the biological and observational errors in the predictor are homoscedastic and uncorrelated.

is, when $\mathbf{V}_u = \sigma_u^2\mathbf{I}$ and $\mathbf{V}_x = \sigma_x^2\mathbf{I}$, then equation (11a) reduces to

$$E[\hat{\beta}_1|\mathbf{x}_o] \approx (1 - \frac{\sigma_u^2}{\sigma_x^2})\beta_1, \tag{12}$$

which we recognize as the standard formula for uncorrelated data with observation error when σ_u^2 and σ_x^2 are estimated as the appropriate sums of squares (e.g., Fuller 1987). We note that this holds even with correlations in the residuals as long as the predictor variables are not correlated. It also holds in the more general situation in which $\mathbf{V}_u = \sigma_u^2\mathbf{A}$ and $\mathbf{V}_x = \sigma_x^2\mathbf{A}$ for a common symmetric positive definite matrix $\mathbf{A}$.

The fact that the bias depends on the correlation structure of the predictor variables shows that the bias cannot be corrected without making assumptions about the predictors' variance structure, which will usually require a model for their evolution.

It is common to quantify the bias from known observation error in the predictor variables with a reliability ratio, K, defined as the ratio between the true and the observed sums of squares for the predictor variable. The true sum of squares is estimated by subtracting the observation variance from the observed sum of squares. For uncorrelated homoscedastic data, $K = 1 - \sigma_u^2/\sigma_x^2$ (Fuller 1987). With correlated data, the corresponding reliability ratio is

$$K \approx 1 - \frac{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{V}_u\mathbf{V}_x^{-1}\mathbf{x}_o}{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o}. \tag{13}$$

In contrast to the standard situation, it is possible to produce negative K, and even $K > 1$ if there are asymmetries in either observational or biological variation across species.

Equations (11a), (11b), and (13) are approximations obtained by assuming that certain averages of the predictor variables are zero (see appendix). If the predictor variables are not centered on their mean, the exact equation (A.6) in the appendix must be used, and then there can also be significant bias in the intercept. We also assume that the relevant variance matrices are known without error. Uncertainty in the variance matrices includes errors in the assumed evolutionary models, in the phylogeny, and in the estimated observation variance, but their effects are beyond the scope of this paper.

In the appendix, we also give general expressions for bias in models with several predictor variables with observation errors that can be correlated both across predictors and species. In general, observation error in one predictor will carry over and cause biases in the coefficients pertaining to other predictors. For uncorrelated data, there is an important exception to this in that observation error in one predictor will not bias the coefficients of another uncorrelated predictor (e.g., Buonaccorsi 2010). In the appendix, we show that this generalizes to models with correlated residuals but not to cases in which the predictor with error is phylogenetically correlated or heteroscedastic. This means that observation errors in random uncorrelated homoscedastic predictors will not bias the coefficients of any fixed effects in the model (asymptotically and provided the predictors are centered on their means). Beyond this, there is not much that can be said about sign and magnitude of the bias with errors in several predictors (Gleser 1992). In the words of Aickin and Ritenbaugh (1996): "there do not seem to be any rules of thumb that would permit one to make even qualitative statements about the nature of this bias." The situation is not less complicated in the presence of phylogenetic correlations.

Correcting the Bias

The bias due to observation error can be corrected by use of the reliability ratio, K, as follows.

$$\hat{\beta}_{1,\text{Corrected}} = \frac{\hat{\beta}_1}{K}, \tag{14}$$

$$\text{Var}[\hat{\beta}_{1,\text{Corrected}}|\mathbf{x}_o] = \frac{1}{K^2\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o}, \tag{15}$$

with K as in equation (13) or more accurately as in equation (A.6) in the appendix. To compute this, we must specify the matrices $\mathbf{V}_e$, $\mathbf{V}_u$, $\mathbf{V}_x$, and $\mathbf{V}=\mathbf{V}_t+\mathbf{V}_e+\beta_1^2\mathbf{V}_{u|x}$.

All these except $\mathbf{V}_x$ are used in the GLS estimation in the first place. The $\mathbf{V}_x$ is the variance matrix of the vector of observed species predictor variables, $\mathbf{x}_o$, which due to the assumed independence of $\mathbf{x}_t$ and $\mathbf{u}$ is $\mathbf{V}_x = \mathbf{V}_{xt} + \mathbf{V}_u$ where $\mathbf{V}_{xt}$ is the true variance matrix of the predictor, which will typically have a phylogenetic structure that must be modeled. For example, assuming that the predictor variables evolve as Brownian motion along the phylogeny would make $\mathbf{V}_{xt}$ a matrix in which the ijth element would be proportional to the shared branch length of species i and species j. The constant of proportionality has to be estimated from the data. If the predictors are treated as fixed effects, $\mathbf{V}_x$ would be their empirical variance matrix. Once all these matrices are specified, the reliability ratio and the bias-corrected estimator can be computed. Numerical simulations reported in online File A (deposited in Dryad, doi:10.5061/dryad.r76cm3bn) indicate that this procedure is effective in removing the bias. An extension to multiple predictors and fixed effects is presented in the appendix.

Although it is tempting to correct a bias, a correction will not always produce a more accurate estimator (online File A). Even when the measurement variance is known, a bias correction can decrease the precision of the estimator, and in addition, there is the error in the reliability ratio itself. In the appendix, we derive a criterion for a bias-corrected slope to be more accurate than an uncorrected slope in the sense of having a lower mean squared error:

$$\frac{\sigma_\beta}{|\beta_1|} < \sqrt{\frac{K^2(1-K)}{1+K}} \quad \text{for } -1 < K < 1, \qquad (16)$$

where $\sigma^2_\beta = 1/(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)$ is the estimated variance of the uncorrected estimator of the slope and K is the reliability ratio as above. For $|K| > 1$, the bias-corrected estimator is always more accurate. The left-hand side of equation (16) is the relative standard error of the uncorrected slope estimate and can be estimated by fitting in an estimator of β_1. As illustrated in Figure 3, for K in the realistic interval between zero and one, it will never pay to correct if the relative error of the slope is above 30%, but if the relative error is less than 10%, bias correction will improve accuracy over a wide range of K (from 0.11 to 0.98).

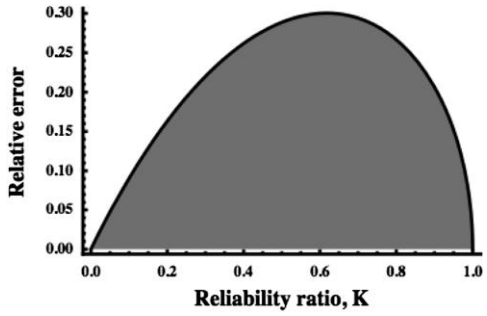

FIGURE 3. Criterion for bias correction to improve accuracy. In the gray area under the curve, the mean squared error of the bias-corrected slope estimate is less than the mean squared error of the uncorrected slope estimate. The relative error refers to the uncorrected slope. This criterion does not include the effects of uncertainty in the estimate of K. Based on equation (16) in the main text.

While equation (16) provides a rule of thumb for when to correct, it should not be the only consideration. The criterion does not include the effects of error in the reliability ratio, which will decrease the accuracy of the bias-corrected estimator, and both the observation variance and the criterion itself are estimated with error. Hence, a conservative approach is advisable, and particularly so when the estimates of the observation variances are unreliable. Bias correction is also not only a statistical consideration. It is more important when slopes with different levels of error are to be compared against each other, and in such situations, it may be advisable to correct bias even if it reduces accuracy.

As will be described elsewhere, we are implementing the above corrections into the program package Slouch (Hansen et al. 2008). This can fit the adaptation and constraint models with observation error based on iterated GLS. In online File B (deposited in Dryad, doi:10.5061/dryad.r76cm3bn), we also provide R-code to implement observation error into standard GLS regression. Since bias correction may not always improve accuracy, we recommend reporting parameter estimates without bias correction and the estimated reliability ratio, so that a possible bias can be evaluated when crucial for argument.

Estimating Observation Variance

In comparative studies based on species means, observation variances for correction will typically be obtained as the squares of the published standard errors of the means. If the sample sizes of individual species are small, such standard errors may be unreliable. In such cases, it might be better to estimate an average across the different species (Ives et al. 2007; Labra et al. 2009; online Files A and B). First, the within-species sample variance is estimated as a sample-size-weighted average of the sample variances of each species; that is, as $\sigma^2_w = \Sigma_i\sigma^2_{wi}(n_i - 1)/\Sigma_i(n_i - 1)$, where σ^2_{wi} is the sample variance of species i and n_i is the sample size of species i. Then, the measurement variance of each species is estimated as σ^2_w/n_i. This procedure assumes that within-species variances are similar across species, and we recommend it for species with sample sizes less than about 20 when this assumption is not obviously violated. Usually, there will be no covariance in observations across species, and the $\mathbf{V}_e$ and $\mathbf{V}_u$ matrices will be diagonal. Kelly and Price (2004) and Felsenstein (2008) present further methods for estimation and incorporation of within-species variation in comparative studies.

BIOLOGICAL ERROR AND PHYLOGENETIC STRUCTURE IN PREDICTOR VARIABLES

Observation error in the predictor variables is not the only way to get a biased estimate of the regression slope.

Just as we may distinguish between observation error and biological error in the model relationship between Y and X, we can make a distinction between observation error and biological error in the predictor variable. Consider a case in which the true biological model is $Y = \beta_0 + \beta_1 Z$, for some variable Z, but we observe not Z, but $X = Z + H$, where H may be considered a biological error. This situation may arise when the relation between Y and X depends only on one component of X. For example, Y may be adult brain size and X adult body size, but adult brain size may depend only on prenatal or juvenile growth and be unaffected by subsequent growth of the organism. Such biological error will have effects that are identical to observation error in the predictor variable but its interpretation is quite different. While the biological error makes us expect a more shallow regression of Y on X, whether this is to be considered a "bias" depends on the exact claim one wants to make. Gould (1975), Lande (1979), and Riska (1989, 1991; Riska and Atchley 1985) have argued for different versions of this mechanism as an explanation for the relatively shallow allometric regressions of brain sizes on body size. We also note that such biological error is likely to be phylogenetically structured and thus suggests a nondiagonal $\mathbf{V}_u$ matrix.

Phylogenetic structure in the predictor variables has consequences that are routinely ignored in evolutionary regression studies. For example, it has become standard practice to check for phylogenetic signal in the response variable so as to decide how to do corrections for phylogeny. This is based on a statistical misunderstanding, as it is the phylogenetic correlations in the model residuals that need be accounted for (e.g., Hansen and Orzack 2005; Labra et al. 2009; Revell 2010). Because the biological variation in the predictor variables may be phylogenetically structured, this will automatically generate a phylogenetic effect in the response variables, even when there is no phylogenetic structure in the model relation (i.e., in the $\mathbf{V}_t$ matrix), and in this situation, a nonphylogenetic regression would be the correct approach. See Labra et al. (2009) for illustration and further discussion of this point.

Alternative Regression Models

The bias that results from observation error in standard least squares regression has led to the popularity of alternative regression models such as reduced major axis regression and structural equation modeling (reviewed in Warton et al. 2006; Smith 2009). These approaches have, however, not been developed to deal properly with correlated residuals, and their application to species data has thus been problematic. This situation has recently improved through the work of Ives et al. (2007). There is, however, a deeper problem with these alternatives in that they do not account for biological error in the regression model. Often, they are motivated by the argument that there is error in both the dependent and the independent variable, and simple ratios of variance in Y and X are used to pick the "right" deviance from the model prediction to minimize. For example, in their influential review of phylogenetic comparative methods, Harvey and Pagel (1991) argued that major axis and reduced major axis regression are more appropriate than standard regression because they allow error in both the X and the Y variables. There is no distinction between biological and observation error in their argument, and their support of structural equations is implicitly based on all the deviance from the model being due to observation error. Above we showed how standard regression models can be adjusted to account for observation error. We will now show that the two most common alternative regression methods are unsuitable when there is biological error in the regression. For convenience, we will ignore phylogenetic correlations (and drop matrix notation, bold face) in this section, as it makes no difference to our argument.

We start with reduced major axis regression. The reduced major axis estimate of a regression slope of Y on X is simply the square root of the ratio of the variances of the two variables (with sign determined by the sign of their correlation). That is,

$$\beta_{\text{rma}} = \sqrt{\frac{\text{Var}[y_o]}{\text{Var}[x_o]}}, \tag{17}$$

which by our model (8) of biological and observation error in both y_o and x_o would be

$$\beta_{\text{rma}} = \sqrt{\frac{\beta_1^2 V_{xt} + V_t + V_e}{V_{xt} + V_u}}. \tag{18}$$

Ignoring estimation error in the variances in equation (17), β_{rma} equals β_1 if $V_t = V_e = V_u = 0$. That is, if the true value of X is the only source of variation in Y, but note that the existence of any variation in Y independent of X or in X independent of Y, be it biological or observational, renders β_{rma} biased (unless $V_t + V_e = \beta_1^2 V_u$, which is clearly coincidental). As illustrated in Figure 4, this bias may be large when V_t is large, which is likely to be the typical situation for evolutionary regressions. Note also that the β_{rma} will typically be large even when β_1 is close to zero. This problem is not solved by Ives et al. (2007) phylogenetic correction of the reduced major axis method based on using phylogenetically weighted sums of squares to estimate $\text{Var}[y_o]$ and $\text{Var}[x_o]$ because these generalized sums of squares can be decomposed as in equation (18).

Model II or major axis regression is another commonly used alternative to standard regression. In this case, the regression slope of Y on X is estimated as the slope of Y on X of the first principal component of the two variables (which must be expressed in the same unit). An estimator of this is (Kuhry and Marcus 1977):

$$\beta_{\text{ma}} = \frac{\text{Var}[y_o] - \text{Var}[x_o] + \sqrt{(\text{Var}[y_o] - \text{Var}[x_o])^2 + 4\text{Cov}[y_o, x_o]^2}}{2\text{Cov}[y_o, x_o]}. \tag{19}$$

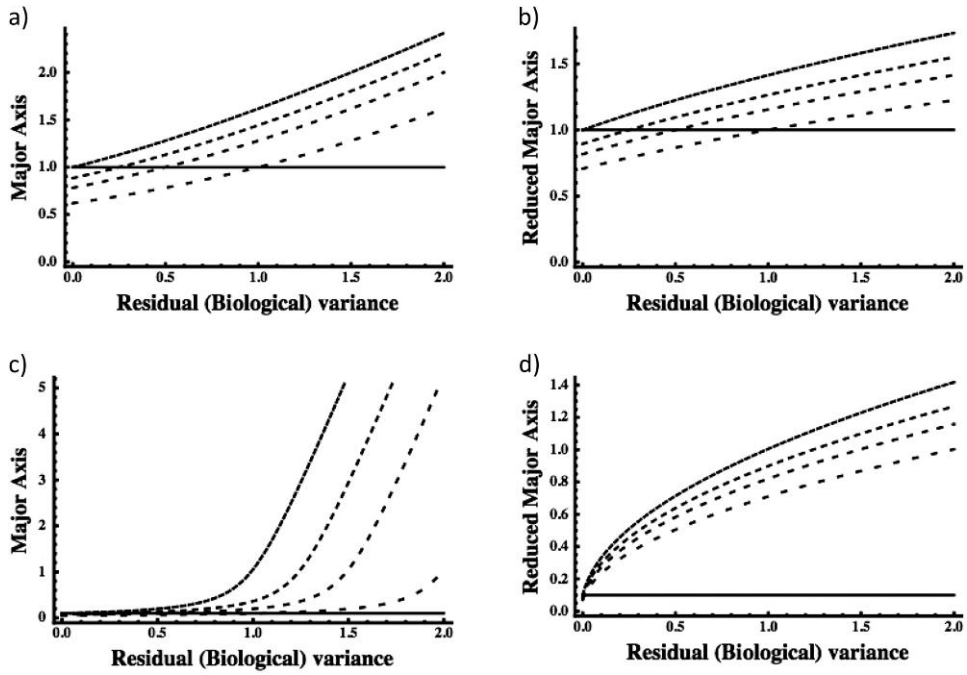

FIGURE 4. Illustration of expected bias in alternative regression models. The expected slopes of, respectively, major axis (a and c) and reduced major axis (b and d) are plotted against the level of residual (biological) variance, V_t, in the model as measured in units of the biological variance in the predictor variable, V_{xt} (i.e., the x-axis is V_t/V_{xt}). In (a) and (b), the true slope (marked by solid line) is 1, and in (c) and (d), it is 0.1. The four dashed lines in each figure illustrate effects of different levels of observation variance, V_u, in the predictor variable, from top to bottom $V_u = 0$, $V_u = 0.25\,V_{xt}$, $V_u = 0.5\,V_{xt}$, and $V_u = V_{xt}$. Observation variance in the dependent variable can be added to V_t. Based on equations (18) and (20) in the main text.

With our model, we have $\mathrm{Cov}[y_o, x_o] = \beta V_{xt}$, $\mathrm{Var}[y_o] = \beta^2 V_{xt} + V_t + V_e$, and $\mathrm{Var}[x_o] = V_{xt} + V_u$. This gives

$$\beta_{\mathrm{ma}} = \frac{(\beta_1^2 - 1)V_{xt} + V_t + V_e - V_u + \sqrt{((\beta_1^2 - 1)V_{xt} + V_t + V_e - V_u)^2 + 4\beta_1^2 V_{xt}^2}}{2\beta_1 V_{xt}}. \quad (20)$$

Ignoring estimation errors of the moments in equation (19), this equals β_1 if $V_t + V_e = V_u$; that is, if the sum of the biological variance around the model and the observation variance in the response variable equals the observation variance in the predictor variable. Because the biological variance, V_t, is likely to be large in most evolutionary regression models, this only seems realistic when V_u also includes biological error in the sense discussed above (i.e., variance in the parameter H). For most realistic combinations of parameters, however, β_{ma} will be a poor estimator of β_1 (Fig. 4). As an example, consider the not unrealistic situation in which measurement variances are zero and $V_t = V_{xt}$, then $\beta_{\mathrm{ma}} = \left(\beta_1 + \sqrt{\beta_1^2 + 4}\right)/2$, which will typically be seriously wrong. In general, there will be an upward bias (steeper slope) as long as biological variance, V_t, is larger than the observation variance of the predictor.

Although this shows that major axis and reduced major axis regression will practically never be good choices for estimating an evolutionary regression slope, a general structured relation model could work by including biological variance in its free variance ratio parameter, λ. The model is (Kuhry and Marcus 1977):

$$\beta_{\mathrm{sr}} = \frac{\mathrm{Var}[y_o] - \lambda \mathrm{Var}[x_o] + \sqrt{(\mathrm{Var}[y_o] - \lambda \mathrm{Var}[x_o])^2 + 4\lambda \mathrm{Cov}[y_o, x_o]^2}}{2\mathrm{Cov}[y_o, x_o]}, \quad (21)$$

which with our assumptions gives

$$\beta_{\mathrm{sr}} = \frac{(\beta_1^2 - \lambda)V_{xt} + V_t + V_e - \lambda V_u + \sqrt{((\beta_1^2 - \lambda)V_{xt} + V_t + V_e - \lambda V_u)^2 + 4\lambda\beta_1^2 V_{xt}^2}}{2\beta_1 V_{xt}}. \quad (22)$$

Choosing the value $\lambda = (V_t + V_e)/V_u$ for the variance ratio parameter would then make β_{sr} a candidate estimator of β_1, although this requires an estimator of V_t, which for an evolutionary regression would depend on process parameters, as discussed above. Ives et al. (2007) showed how to incorporate known among-species correlations in these models but did not consider biological variance in their implementation of λ.

DISCUSSION

Our analysis reveals and quantifies three distinct sources of bias in the regression slopes estimated from

the phylogenetic comparative method based on GLS. The first is due to a difference between the optimal relation between the variables and the observed evolutionary regression, which will be more shallow if the species have not had time to adapt perfectly to the optimal relation. This can only be considered a statistical bias if the goal is to estimate the evolutionary optimal relation between the variables. It is not a statistical bias in the evolutionary regression per se. This bias particularly matters if the goal is to test a prediction from an optimality model or if a comparison is to be made between different groups that may have had different amounts of time available for adaptation (e.g., different taxonomic levels). In such situations, the bias can be quantified and corrected by use of the phylogenetic correction factor, ρ, introduced by Hansen et al. (2008). See also Burt (1989) and Deaner and Nunn (1999) for other ways of quantifying this bias. The second bias is due to observation error in the predictor variables, which typically leads to a more shallow regression slope. This bias is important if observation error constitutes a significant component of the total variation in the predictor variables. Although a formal correction for this bias, known as the reliability ratio or attenuation factor, K, has long been available in the statistical literature (e.g., Fuller 1987), it has rarely been used in cross-species regression studies (Taper and Marquet 1996; Borrell 2007 are two exceptions). In this paper, we have derived a general phylogenetically corrected reliability ratio, which can be used to assess the bias resulting from known observation variance provided one has an explicit model for the true variation in both the response and predictor variables. The third bias is due to biological "error" in the predictor variables, which occurs when only a component of the observed predictor variable is causally related to the response variable. If the predictor can be decomposed into its causal and noncausal components, this bias can also be corrected by treating the noncausal variance as measurement variance in the phylogenetic reliability ratio. Even if such decomposition may rarely be possible without direct knowledge of the parts, biological error in the predictor has important implications for the interpretation of the evolutionary regression, and our formalization may help draw attention to this possibility.

It is illustrative how all these sources of bias have appeared as possible explanations of the observation that evolutionary regression slopes are often more shallow on lower than on higher taxonomic levels (e.g., Martin and Harvey 1985). Explanations based on evolutionary lag, or maladaptation, being more pronounced at lower taxonomic levels have been suggested by, for example, Burt (1989), Deaner and Nunn (1999), and Hansen et al. (2008), explanations based on biological error in the predictor being relatively more important on low taxonomic levels have been suggested by, for example, Lande (1979) and Riska (1991), and explanations based on observation errors being relatively more important on lower taxonomic levels have been suggested by, for example, Pagel and Harvey (1988, 1989). What combination of these mutually consistent explanations is correct is an empirical question that must be assessed on a case-by-case basis.

Evolutionary regressions can be generated by a variety of mechanisms. We have outlined two generic mechanisms representing adaptation and constraint. The main difference between these is that the adaptation model has a necessary link between correlations in residual deviations and bias in the estimated slope, whereas this link, and indeed the bias, is absent in the constraint model. The adaptation model also predicts some form of exponential decay of the residual covariances. This is possible, but not necessary, under the constraint model, which may also be consistent with residual covariances that depend only on shared branch lengths as under Brownian motion. The choice between these models should be based on biological more than statistical considerations. If one wants to test an optimality model or functionality prediction, one should use the adaptation model. If one wants to test a constraint hypothesis or simply correct for allometric influence of size on a trait, then the constraint model is appropriate.

We are far from the first to discuss how observation error or within-species sampling error can be included in the phylogenetic comparative methods. Various explicit methods for doing so have been presented by Lynch (1991), Martins and Hansen (1997), Housworth et al. (2004), Ives et al. (2007), Adams (2008), Felsenstein (2008), Lajeunesse (2009), and Hadfield and Nakagawa (2010), and there are numerous discussions and evaluations of the problem (e.g., Pagel and Harvey 1988; Kelly and Price 2004, 2005; Harmon and Losos 2005). Of these, Ives et al. (2007) present the most detailed and general approaches, including methods for incorporating observation error into general structural equations and phylogenetic effect estimation. Despite this, we perceive a disconnect between discussion of statistical methods on one hand and conceptual discussion as to the biological meaning of residual deviations on the other. Our aim in this paper has been to describe the complex distinction and interplay between biological and observation error as explicitly as possible and also to present formulas to quantify the various biases in the phylogenetic regression. Note also that our corrections for observation error are not restricted to comparative studies, but are completely general, and can be used with any GLS regression model regardless of the source of observation error. They do not apply to evolutionary correlation coefficients, however, which are generally biased by observation error in both variables (Adolph and Hardin 2007; Felsenstein 2008).

This approach presupposes that an a priori estimate of the observation variance is available. In comparative studies based on means, this will typically be based on a standard error of the mean. Arguably, this will often underestimate the observation variance, which could also include error due to various forms of nonrandom sampling of individuals and population structure, but this does not mean that what is available should not be used.

We warn again, however, that standard errors computed from small samples are unreliable and should not be used uncritically. We also note that there is a considerable literature on how to deal with observation error when the observation variances are not directly known. This is all based on some form of incomplete information, for example, assumptions about the ratio of the observation variances of the dependent and independent variables (Fuller 1987). Most of these approaches will run into problems if there is biological error in the model, and it seems to us that the best strategy in practice will be to obtain direct estimates of the observation variances.

It should be clear from our analysis that major axis and reduced major axis regression are not acceptable estimators of evolutionary regression slopes when there is biological error in the model. Still, these methods, and reduced major axis in particular, are so common that their users often do not even bother to give a justification (Smith 2009). This situation remains in the face of regular criticisms (e.g., Seim and Sæther 1983; Kelly and Price 2004; Smith 2009). Kelly and Price (2004) show by example how both these methods give results that are far away from maximum-likelihood estimates in three biological systems in which biological error is present.

Our analysis indicates that reduced major axis slopes can only be reliable when we have a near perfect relation between the two variables; that is, when neither biological nor observation errors are present. With any form of error, a correspondence between the reduced major axis and an underlying evolutionary regression will be coincidental. In our opinion, reduced major axis regression should never have been used to estimate evolutionary or allometric regression slopes, and every published result based on this method needs to be reconsidered.

Major axis regression is in much the same situation and can only be considered reasonable as a rough estimate when the residual variation of the model is much less than the true predictor variation. General structured relations can be used if biological variance is properly included in the variance ratio parameter, but we see no advantage over standard phylogenetic regression analysis with bias correction.

This does not mean that these methods are not acceptable as estimators of other parameters. The reduced major axis may be a sensible estimator of the ratio between the standard deviations of two variables. The major axis regression should be regarded as an estimator of the slope of the first principal component of the two variables. The first principal component is an important descriptor of the relationship between variables, and this shows that major axis regression has an important role to play. The problem is simply that the slope of the major axis is not the same parameter as the regression slope represented in our model (1). We thus want to rediagnose the problem from "what is the best estimator?" to "what parameters are we estimating?" In this perspective, it is clear that the alternative regression methods are not statistical alternatives that should be used when there are particular patterns of error variance, but biological alternatives that should be used when there are different biological questions (cmp. Houle et al. 2011). Warton et al. (2006) and Smith (2009) diagnosed the situation similarly, arguing that the choice of regression methods should be guided by biological and not statistical considerations, but we disagree with their specific recommendations for doing so. Smith's (2009) recommendations for the use of reduced major axis seem for the most part to call for major axis regression, which he did not consider, and the recommendations in Warton et al. (2006) are not based on explicit models of biological error and are not supported by the results in Kelly and Price (2004) and the present paper.

In summary, when the biological goal is to estimate a linear causal effect of a predictor variable X on Y, as for example, when we study adaptation of a biological trait Y to an environmental variable X, then standard (GLS) regression methods should be used. Alternative major axis and reduced major axis regressions should never be used in this situation regardless of the pattern and type of error variation in the variables. This also extends to the standard allometric regressions of Log Y on Log X, where X is body size, when the goal is to estimate the allometric exponent, b, in the equation $Y = aX^b$. The presence of either biological or observational error in the variables of these regressions is indeed a statistical problem that should be accounted for, but this should be done within the framework of the phylogenetic GLS regression, and the methods presented in this paper provide the basic tools for doing so.

Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found at http://datadryad.org (doi:10.5061/dryad.r76cm3bn).

Funding

K.B. was supported by the Centre for Theoretical Biology at University of Gothenburg.

Acknowledgments

We thank David Houle, Christophe Pélabon, and Jason Pienaar for discussions and particularly Jason Pienaar for help in implementing the methods in Slouch.

References

Adams D.C. 2008. Phylogenetic meta-analysis. Evolution. 62:567–572.
Adolph S.C., Hardin J.S. 2007. Estimating phenotypic correlations: correcting for bias due to intraindividual variability. Funct. Ecol. 21:178–184.
Aickin M., Ritenbaugh C. 1996. Analysis of multivariate reliability structures and the induced bias in linear model estimation. Stat. Med. 16:1647–1661.

Borrell B.J. 2007. Scaling of nectar foraging in orchid bees. Am. Nat. 169:569–580.
Buonaccorsi J.P. 2010. Measurement error: models, methods, and applications. New York: Chapman & Hall.
Burt A. 1989. Comparative methods using phylogenetically independent contrasts. Oxf. Surv. Evol. Biol. 6:33–53.
Butler M.A., King A.A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683–695.
Deaner R.O., Nunn C.L. 1999. How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. Proc. R. Soc. Lond. B. Biol. Sci. 266:687–694.
Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 126:1–25.
Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. Am. Nat. 171:713–725.
Fuller W.A. 1987. Measurement error models. Hoboken (NJ): Wiley.
Fuller W.A. 1995. Estimation in the presence of measurement error. Int. Stat. Rev. 63:121–147.
Gleser L.J. 1992. The importance of assessing measurement reliability in multivariate regression. J. Am. Stat. Assoc. 87:696–707.
Gould S.J. 1975. Allometry in primates with emphasis on scaling and the evolution of the brain. Contrib. Primatol. 5:244–292.
Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 326:119–157.
Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494–508.
Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation . Evolution. 51:1341–1351.
Hansen T.F., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution. 50:1404–1417.
Hansen T.F., Orzack S.H. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution. 59:2063–2072.
Hansen T.F., Pienaar J., Orzack S.H. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution. 62:1965–1977.
Harmon L.J., Losos J.B. 2005. The effect of intraspecific sample size on type I and type II error rates in comparative studies. Evolution. 59:2705–2710.
Harvey P.H., Pagel M.D. 1991. The comparative method in evolutionary biology. Oxford: Oxford University press.
Houle D., Pélabon C., Wagner G.P., Hansen T.F. 2011. Measurement and meaning in biology. Q. Rev. Biol. 86:3–34.
Housworth E.A., Martins E.P., Lynch M. 2004. The phylogenetic mixed model. Am. Nat. 163:84–96.
Huxley J.S. 1932. Problems of relative growth. Baltimore (MD): The John Hopkins University Press. Reprint in 1993 with contributions of F.B. Churchill and R.E. Strauss.
Ives A.R., Midford P.E., Garland T. Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56:252–270.
Kelly C., Price T.D. 2004. Comparative methods based on species mean values. Math. Biosci. 187:135–154.
Kelly C., Price T.D. 2005. Correcting for regression to the mean. Am. Nat. 166:700–707.
Kuhry B., Marcus L.F. 1977. Bivariate linear models in biometry. Syst. Zool. 26:201–209.
Labra A., Pienaar J., Hansen T.F. 2009. Evolution of thermal physiology in *Liolaemus* lizards: adaptation, phylogenetic inertia, and niche tracking. Am. Nat. 174:204–220.
Lajeunesse M.J. 2009. Meta-analysis and the comparative phylogenetic method. Am. Nat. 174:369–381.
Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution. 33:402–416.
Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution. 45:1065–1080.
Martin R.D., Harvey P.H. 1985. Brain size allometry: ontogeny and phylogeny. In: Jungers W.L., editor. Brain size allometry: ontogeny and phylogeny. New York: Plenum.
Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149: 646–667.
Nee S., Read A.F., Harvey, P.H. 1996. Why phylogenies are necessary for comparative analysis. In Martins E.P., editor. Phylogenies and the comparative method in animal behavior. Oxford: Oxford University Press. p. 399–411.
Pagel M.D., Harvey P.H. 1988. The taxon-level problem in the evolution of mammalian brain size: facts and artifacts. Am. Nat. 132: 344–359.
Pagel M.D., Harvey P.H. 1989. Taxonomic differences in the scaling of brain on body size among mammals. Science. 244:1589–1593.
Price T. 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 352:519–529.
Revell L.J. 2010. Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. 1:319–329.
Ridley M. 1989. Why not to use species in comparative tests. J. Theor. Biol. 136:361–364.
Riska B. 1989. Composite traits, selection response, and evolution. Evolution. 43:1172–1191.
Riska B. 1991. Regression models in evolutionary allometry. Am. Nat. 138:283–299.
Riska B., Atchley W.R. 1985. Genetics of growth predict patterns of brain-size evolution. Science. 229:668–671.
Savageau M.A. 1979. Allometric morphogenesis of complex systems: derivation of the basic equations from first principles. Proc. Natl. Acad. Sci. U S A. 76:6023–6025.
Seim E., Sæther B.-E. 1983. On rethinking allometry: which regression model to use? J. Theor. Biol. 104:161–168.
Smith R.J. 2009. Use and misuse of reduced major axis for line-fitting. Am. J. Phys. Anthropol. 140:476–486.
Taper M.L., Marquet P.A. 1996. How do species really divide resources? Am. Nat. 147:1072–1086.
Warton D.I., Wright I.J., Falster D.S., Westoby M. 2006. Bivariate line-fitting methods for allometry. Biol. Rev. 81:259–291.

APPENDIX

Consider the model

$$\mathbf{y} = \mathbf{y}_t + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e),$$
$$\mathbf{D} = \mathbf{D}_t + \mathbf{U}, \mathrm{vec}[\mathbf{D}_t] \sim N(\mathbf{a}, \mathbf{V}_d), \mathrm{vec}[\mathbf{U}] \sim N(\mathbf{0}, \mathbf{V}_u),$$
$$\mathbf{y}_t = \mathbf{D}_t \boldsymbol{\beta} + \mathbf{r}_t, \mathbf{r}_t \sim N(\mathbf{0}, \mathbf{V}_t), \quad \text{(A.1)}$$

where $\mathbf{y}$ is a vector of n observations of the dependent variable, $\mathbf{D}$ is an observed $n \times m$ design matrix, $\mathbf{e}$ is a vector of n observation errors in the dependent variable, $\mathbf{U}$ is an $n \times m$ matrix of observation errors in the elements of $\mathbf{D}$, $\mathbf{y}_t$ is an vector of n true values of $\mathbf{y}$, $\mathbf{D}_t$ is an $n \times m$ matrix of true values of $\mathbf{D}$, $\boldsymbol{\beta}$ is a vector of m true values of the parameters to be estimated, $\mathbf{r}_t$ is a vector of n residual deviations in the true model, and the vec operator forms a vector by stacking the columns of its matrix argument. We assume that $\mathbf{e}$, $\mathbf{D}_t$, $\mathbf{U}$, and $\mathbf{r}_t$ are independent of each other and that all variance matrices are positive definite, except that $\mathbf{V}_d$ and $\mathbf{V}_u$ may be block matrices with zero blocks corresponding to fixed effects and variables without measurement error, respectively.

To estimate the parameter vector, we use the model $\mathbf{y} = \mathbf{D}\hat{\boldsymbol{\beta}} + \mathbf{r}$, where the residual vector, $\mathbf{r}$, is assumed to have a variance matrix $\mathbf{V} = \mathbf{V}_t + \mathbf{V}_e + \mathrm{Var}[\mathbf{U}\boldsymbol{\beta}|\mathbf{D}] = \mathbf{V}_t + \mathbf{V}_e + \Sigma_i \Sigma_j \beta_i \beta_j \mathbf{V}_{uij}$, where $\mathbf{V}_{uij}$ is the appropriate submatrix of $\mathbf{V}_u - \mathbf{V}_u (\mathbf{V}_d + \mathbf{V}_u)^- \mathbf{V}_u$ that contains the between

species covariances in the observation errors of predictors i and j. Conditional on $\mathbf{V}$ and $\mathbf{D}$, the naive GLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}\mathbf{y} = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{D}_t\boldsymbol{\beta} + \mathbf{r}_t + \mathbf{e})$$
$$= \boldsymbol{\beta} + (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}(-\mathbf{U}\boldsymbol{\beta} + \mathbf{r}_t + \mathbf{e}). \quad (A.2)$$

The expectation of this conditional on the observed values of the predictor variables is

$$E[\hat{\boldsymbol{\beta}}|\mathbf{D}] = (\mathbf{I} - (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}E[\mathbf{U}|\mathbf{D}])\boldsymbol{\beta}, \quad (A.3)$$

where $\mathbf{I}$ is an $m \times m$ identity matrix. It remains to calculate E[$\mathbf{U}$|$\mathbf{D}$]. To do this, we use the vectorized form and the assumptions of normal distributions and independence

$$E[\mathrm{vec}[\mathbf{U}]|\mathbf{D}] = \mathrm{Cov}[\mathrm{vec}[\mathbf{U}], \mathrm{vec}[\mathbf{D}]]\mathrm{Var}[\mathrm{vec}[\mathbf{D}]]^{-}(vec[\mathbf{D}] - \mathbf{a})$$
$$= \mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-}(\mathrm{vec}[\mathbf{D}] - \mathbf{a}), \quad (A.4)$$

where we have used $\mathrm{Cov}[\mathrm{vec}[\mathbf{U}], \mathrm{vec}[\mathbf{D}]] = \mathrm{Cov}[\mathrm{vec}[\mathbf{U}], \mathrm{vec}[\mathbf{U}] + \mathrm{vec}[\mathbf{D}_t]] = \mathrm{Var}[\mathrm{vec}[\mathbf{U}]] = \mathbf{V}_u$. Also note that we have used generalized inverse (denoted by "$^-$") to account for fixed effects without biological and observational variance in $\mathbf{V}_u$ and $\mathbf{V}_d$. Using equation (A.4) in equation (A.3) gives

$$E[\hat{\boldsymbol{\beta}}|\mathbf{D}] = \mathbf{K}\boldsymbol{\beta} = (\mathbf{I} - (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}$$
$$\times \mathrm{vec}^{-1}[\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-}(\mathrm{vec}[\mathbf{D}] - \mathbf{a})])\boldsymbol{\beta}, \quad (A.5)$$

where vec^{-1} denotes the inverse of the vec operator. To use this, we need to estimate the vector $\mathbf{a}$, which contains elements of two types. One type is simply the values of fixed effects in $\mathbf{D}$, and the corresponding elements in vec[$\mathbf{D}$] $-$ $\mathbf{a}$ are zero. The other type is means of the random effects in the model, and like the various variance matrices in the model, these need be estimated. Most naturally, this is done as a GLS mean over the predictor variables.

Equation (A.5) is the basis for equations (11a, 11b to 13) in the main text. Consider first the case of a model with an intercept, β_0, and a single predictor, $\mathbf{x}_o$, with parameter β_1. This gives

where the $\mathbf{V}_d$ and $\mathbf{V}_u$ here are to be interpreted as the nonzero submatrices pertaining to the predictor variable and its error and $<\mathbf{x}_o>$ is the expected value of the predictor vector, which can be estimated as $<x_o>\mathbf{1}$ with $< x_o > = \mathbf{x}_o^T(\mathbf{V}_d + \mathbf{V}_u)^{-1}\mathbf{1}/\mathbf{1}^T(\mathbf{V}_d + \mathbf{V}_u)^{-1}\mathbf{1}$. These equations are unwieldy looking, but can be greatly simplified by assuming that the predictor variables have mean zero, and making the approximation $\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{1} \approx 0$. This gives

$$E\left[\hat{\beta}_0\,|\mathbf{x}_o\right] \approx \beta_0 - \frac{\mathbf{1}^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}\mathbf{x}_o}{\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}}\beta_1, \quad (A.7a)$$

$$E\left[\hat{\beta}_1\,|\mathbf{x}_o\right] \approx K\beta_1 \approx \left(1 - \frac{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}\mathbf{x}_o}{\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o}\right)\beta_1. \quad (A.7b)$$

With centered predictor variables, the bias in the intercept would normally be small for large data sets, but this is not true if the predictor variables are not centered, in which case we have to use equation (A.6) above to get a reasonable correction. This result generalizes to the case with several (error-free) fixed effects and a single random predictor variable. If the predictor variable is centered on its mean, the estimates of the fixed effects should have small bias, and the bias in the coefficient of the random predictor is approximated by (A.7b). From equation (A.7b), we note that observation error usually makes the estimated slope too shallow, but, since the matrix $\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}$ does not need to be positive definite, it is possible to make $K > 1$ and also $K < 0$ in certain situations. This may occur when variation is very asymmetric across species.

Note for completeness that centering of predictor variables would remove one degree of freedom and cause a correlation that ideally should be included in the relevant variance matrices (and require a generalized inverse in equation (A.7) to account for the resulting singularity). In practice, however, the error of ignoring this would be small for a reasonable number of species.

The situation with measurement error in several predictors is much more complicated. To aid intuition, we present a few results on special cases. First, if each of the predictor variables are biologically and observationally uncorrelated across species, and each identically distributed across species (with mean zero) such that $\mathbf{V}_d = \boldsymbol{\Sigma}_d \otimes \mathbf{I}$ and $\mathbf{V}_u = \boldsymbol{\Sigma}_u \otimes \mathbf{I}$, where $\mathbf{I}$ is an $n \times n$ identity

$$E\left[\hat{\beta}_0\,|\mathbf{x}_o\right] = \beta_0 - \frac{(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}(\mathbf{x}_o - < \mathbf{x}_o >)) - (\mathbf{1}^T\mathbf{V}^{-1}\mathbf{x}_o)(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}(\mathbf{x}_o - < \mathbf{x}_o >))}{(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}) - (\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{1})^2}\beta_1, \quad (A.6a)$$

$$E\left[\hat{\beta}_0\,|\mathbf{x_o}\right] = K\beta_1 = \left(1 - \frac{(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}(\mathbf{x}_o - < \mathbf{x}_o >)) - (\mathbf{1}^T\mathbf{V}^{-1}\mathbf{x}_o)(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{V}_u(\mathbf{V}_d + \mathbf{V}_u)^{-1}(\mathbf{x}_o - < \mathbf{x}_o >))}{(\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}) - (\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{1})^2}\right)\beta_1, \quad (A.6b)$$

matrix and $\mathbf{\Sigma}_d$ and $\mathbf{\Sigma}_u$ are $m \times m$ variance matrices with parameters describing the biological and observational covariances between the m predictors, then

$$E[\hat{\boldsymbol{\beta}}|\mathbf{D}] = (\mathbf{I} - (\mathbf{\Sigma}_d + \mathbf{\Sigma}_u)^{-1}\mathbf{\Sigma}_u)\boldsymbol{\beta}, \tag{A.8}$$

where $\mathbf{I}$ is an $m \times m$ identity matrix. This generalizes the standard equation (12) to multiple predictors. This formula is well known for independent data (e.g., Buonaccorsi 2010, p. 109), and our derivation shows that it also holds with a nondiagonal $\mathbf{V}$ matrix (only the predictors need be phylogenetically uncorrelated, not the response variables). The matrix $\hat{\mathbf{\Sigma}}_u$ contains the estimated observation variances and covariances of the predictors, and the matrix $\mathbf{\Sigma}_d + \mathbf{\Sigma}_u$ is the observed variance matrix of the predictors. This equation shows that measurement error in one predictor variable will typically bias the coefficients of correlated predictor variables but not those of uncorrelated predictors. The latter result does not generalize to phylogenetically correlated predictors, however. To see this, let us assume that there are no fixed effects in the model, and let $\mathbf{V}_d$ and $\mathbf{V}_u$ be block diagonal with $mn \times n$ diagonal blocks $\mathbf{V}_{dii}$ and $\mathbf{V}_{uii}$, respectively. With these assumptions, equation (A.5) becomes

$$E[\hat{\boldsymbol{\beta}}|\mathbf{D}] = (\mathbf{I} - (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1}[\mathbf{V}_{u11}(\mathbf{V}_{d11} + \mathbf{V}_{u11})^{-1}\mathbf{d}_1, \ldots, \mathbf{V}_{umm}(\mathbf{V}_{dmm} + \mathbf{V}_{umm})^{-1}\mathbf{d}_m])\boldsymbol{\beta}, \tag{A.9}$$

where $\mathbf{d}_i$ is the ith column of the $\mathbf{D}$ matrix. Because the $\mathbf{V}$ matrix contains observation variance from all variables, this shows that observation variance in one predictor will affect the coefficients of other predictors with the exception that the coefficient of a predictor that is itself observed without error will not be biased by observation error in other predictors that are phylogenetically uncorrelated and homoscedastic (if $\mathbf{V}_{uii}$ is zero, the coefficient β_i will be unbiased if $\mathbf{V}_{ujj}$ and $\mathbf{V}_{djj}$ are both proportional to the identity matrix for all j).

The variance matrix of the naive GLS estimator is $\mathrm{Var}[\hat{\boldsymbol{\beta}}|\mathbf{D}] = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}$, and conditionally on $\mathbf{K}$, the variance of the bias-corrected estimator $\mathbf{K}^{-1}\hat{\boldsymbol{\beta}}$ is $\mathrm{Var}[\mathbf{K}^{-1}\hat{\boldsymbol{\beta}}|\mathbf{D}] = \mathbf{K}^{-1}(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{K}^{-T}$. For the special case of the single bias-corrected slope, $\hat{\beta}_{1,\mathrm{Corrected}} = \hat{\beta}_1/K$, the variance is $\mathrm{Var}[\hat{\beta}_{\mathrm{Corrected}}|\mathbf{x}_0] = 1/(K^2((\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)^{-1}))$. With $|K| < 1$, this variance is larger than that of an uncorrected slope, and then the bias-corrected slope is less precise than an uncorrected slope, and also not maximum likelihood. It is thus not given that the overall accuracy is improved by bias correction (see online File A). To explore when bias correction is likely to increase accuracy, we can compute the mean squared errors of corrected and uncorrected slopes. Since the mean squared error is equal to variance plus squared bias, we have

$$\sigma^2_\beta/K^2 < \sigma^2_\beta + (K-1)^2\beta_1^2, \tag{A.10}$$

where $\sigma^2_\beta = (\mathbf{x}_o^T\mathbf{V}^{-1}\mathbf{x}_o)^{-1}$, as a criterion for the mean squared error of the bias-corrected estimator to be less than the mean squared error of the uncorrected estimator. We can immediately see that this is always fulfilled if $|K| > 1$, and if $|K| < 1$, we get equation (16) in the main text. Similar results on the accuracy of bias correction with uncorrelated data can be found in Gleser (1992) and Fuller (1995), who also consider multiple predictors.