Independent Contrasts and PGLS Regression Estimators Are Equivalent

SIMON P. BLOMBERG[1,*], JAMES G. LEFEVRE[2], JESSIE A. WELLS[1], AND MARY WATERHOUSE[3]

[1]*School of Biological Sciences, University of Queensland, St. Lucia, Queensland 4072, Australia;* [2]*Centre for Discrete Mathematics and Computing,
University of Queensland, St. Lucia, Queensland 4072, Australia; and* [3]*Discipline of Mathematical Sciences, Faculty of Science and Technology,
School of Mathematical Sciences, Queensland University of Technology,
Brisbane, Queensland 4001, Australia;*
*Correspondence to be sent to: School of Biological Sciences, University of Queensland, St. Lucia, Queensland 4072, Australia;
E-mail: s.blomberg1@uq.edu.au.*

Abstract.—We prove that the slope parameter of the ordinary least squares regression of phylogenetically independent contrasts (PICs) conducted through the origin is identical to the slope parameter of the method of generalized least squares (GLSs) regression under a Brownian motion model of evolution. This equivalence has several implications: 1. Understanding the structure of the linear model for GLS regression provides insight into when and why phylogeny is important in comparative studies. 2. The limitations of the PIC regression analysis are the same as the limitations of the GLS model. In particular, phylogenetic covariance applies only to the response variable in the regression and the explanatory variable should be regarded as fixed. Calculation of PICs for explanatory variables should be treated as a mathematical idiosyncrasy of the PIC regression algorithm. 3. Since the GLS estimator is the best linear unbiased estimator (BLUE), the slope parameter estimated using PICs is also BLUE. 4. If the slope is estimated using different branch lengths for the explanatory and response variables in the PIC algorithm, the estimator is no longer the BLUE, so this is not recommended. Finally, we discuss whether or not and how to accommodate phylogenetic covariance in regression analyses, particularly in relation to the problem of phylogenetic uncertainty. This discussion is from both frequentist and Bayesian perspectives. [Comparative methods; generalized least squares; independent contrasts; regression.]

Phylogenetic comparative methods provide key tools for studies of trait evolution and have been applied in many studies of relationships of species to their environments and correlated evolution of traits among species (Harvey and Pagel 1991; Garland et al. 1999). Felsenstein's (1985) phylogenetically independent contrasts (PIC) method has been widely applied in the literature to analyze cross-species data sets by linear regression and correlation methods. It has probably been too successful in that the enthusiastic application of PICs has greatly outpaced our theoretical understanding of the method and its relations to the rest of statistical theory. Several authors have commented that the PIC procedure produces identical results to a generalized least squares (GLS) linear model if one assumes a Brownian motion model of evolution for the traits in question in the latter (see Garland and Ives 2000; Rohlf 2001). One further disadvantage is that the PIC procedure does not automatically give an estimate of the intercept β_0, but this can be calculated (Garland et al. 1993; Garland and Ives 2000). The purpose of this paper is to clarify the relationship between PIC and GLS procedures, and hence to elucidate the use of phylogenies for regression analyses in general, with particular attention to the case where the phylogeny may be in error.

It is clear that phylogenetic covariance is very common in cross-species data (Freckleton et al. 2002; Blomberg et al. 2003). This means that phylogenetic covariance must be considered in analyses of species traits (via regression or correlational methods) if the aim is to obtain unbiased and minimum variance estimators of cross-species relationships among variables. This is true whether or not the aim is to make inferences on adaptation or trait coadaptation. Early developers of phylogenetic comparative methods emphasized the study

of adaptation; however, there is still contention over whether it is possible to use phylogenetic comparative methods to study processes of adaptation (Leroi et al. 1994; Doughty 1996; Martins 2000; Kluge 2005).

Phylogenetic covariance is a property of the data not (necessarily) a property of the scientific hypotheses under investigation. Thus, phylogenetic comparative methods are similar in many ways to the analysis of time-series or geostatistical data (Ives and Zhu 2006). All 3 data types challenge the researcher to deal with data that break the usual statistical assumption of independence. Here, we provide a proof that the PIC method is formally identical to phylogenetic generalized least squares (PGLS). We discuss the frequentist properties of the resulting regression estimators and suggest a Bayesian approach.

GLS VERSUS PIC

Here and throughout the rest of the paper, we discuss GLS and PIC regressions that are bivariate only, however, the results generalize to multiple regression quite easily. The proof of the equivalence of GLS and PIC regression has been described as "awkward" (Rohlf 2001), however, the result is well known among comparative methods researchers. Grafen (1989) provides a proof of a similar result. Garland and Ives (2000) also provide an indication of the equivalence of GLS and PIC estimators. We provide a new proof in Appendix 1. Our proof is notationally clear, relying only on some simple matrix algebra. We prove that $\hat{\beta}_{GLS} = \hat{\beta}_{PIC}$ for the simplest case: one tip species and one ancestor. We then show by strong induction (e.g., Seroul 2000) that the result holds for any arbitrary tree. As a demonstration

for readers who do not find the proof accessible, the simple case of a 2-species tree is given below with intermediate steps. Such a demonstration is pedagogically useful when examining the similarities and differences between PIC and GLS estimators. All the mathematics was performed using the Axiom computer algebra system (Jenks and Sutor 1992).

Consider 2 species a and b joined to a root node with branches v_1 and v_2, respectively. Species a has trait values (x_a, y_a) and species b has trait values (x_b, y_b). We are interested in the regression of y on x, accounting for phylogenetic covariance. Using the PIC algorithm, only one contrast exists for the species a versus species b comparison. However, we can still fit the regression because it is forced through the origin. Since there are only 2 points (including the origin), the residual variance $\sigma^2 = 0$. The value for the contrast for each variable x and y is just $\frac{x_b - x_a}{\sqrt{v_2 + v_1}}$ and $\frac{y_b - y_a}{\sqrt{v_2 + v_1}}$. Now, the regression of $\mathbf{y}$ PICs on $\mathbf{X}$ PICs through the origin is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'_{\mathbf{PIC}}\mathbf{X}_{\mathbf{PIC}})^{-1}\mathbf{X}'_{\mathbf{PIC}}\mathbf{y}_{\mathbf{PIC}}, \tag{1}$$

where $\mathbf{X}_{\mathbf{PIC}}$ and $\mathbf{y}_{\mathbf{PIC}}$ are vectors of the independent contrasts calculated for the explanatory and response variables, respectively. We use uppercase $\mathbf{X}$ for the explanatory variable to be consistent with the terminology for GLS, as $\mathbf{X}_{\mathbf{GLS}}$ is a matrix, whereas $\mathbf{X}_{\mathbf{PIC}}$ is a column vector. The "'" refers to the transpose of a vector, in this case $\mathbf{X}'_{\mathbf{PIC}}$ means $\mathbf{X}_{\mathbf{PIC}}$ is transposed from a column vector to become a row vector. Note that for the two-taxon case, $\mathbf{X}_{\mathbf{PIC}}$ and $\mathbf{y}_{\mathbf{PIC}}$ both have a length of 1.

Substituting the contrasts into Equation (1), we obtain

$$\hat{\boldsymbol{\beta}}_{\mathbf{PIC}} = \left(\frac{x_b - \mathbf{x}_a}{\sqrt{v_1 + v_2}} \frac{x_b - x_a}{\sqrt{v_1 + v_2}} \right)^{-1} \left(\frac{x_b - x_a}{\sqrt{v_1 + v_2}} \right)' \left(\frac{y_b - y_a}{\sqrt{v_1 + v_2}} \right)$$

$$= \left(\frac{v_2 + v_1}{(x_b - x_a)^2} \right) \left(\frac{x_b - x_a}{\sqrt{v_1 + v_2}} \right) \left(\frac{y_b - y_a}{\sqrt{v_1 + v_2}} \right)$$

$$= \frac{y_b - y_a}{x_b - x_a}, \tag{2}$$

which is also just the ordinary least squares (OLS) estimator in this case.

The GLS linear model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{MVN}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}), \tag{3}$$

where $\mathbf{X}$ is a two-column matrix with ones in the first column to account for the intercept. Values for the explanatory variable are in the second column of $\mathbf{X}$. $\mathbf{y}$ is a column vector and $\boldsymbol{\Sigma}$ is the matrix describing the evolutionary covariance of the data due to common ancestry. $\boldsymbol{\varepsilon}$ is a column vector of errors, multivariate normally distributed, with scaling constant σ^2 and with a mean of 0. $\boldsymbol{\beta}$ is a column vector of parameters, of length 2 (intercept and slope). The GLS estimator is

$$\hat{\boldsymbol{\beta}}_{\mathbf{GLS}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{4}$$

Equivalently, we can transform Equation (3) to achieve independent errors by premultiplying by $\boldsymbol{\Sigma}^{\frac{-1}{2}}$ to form $\mathbf{X}^*$ and $\mathbf{y}^*$ and then use the OLS estimator on the transformed data to obtain the GLS estimates for $\hat{\beta}$:

$$\hat{\boldsymbol{\beta}}_{\mathbf{GLS}} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^*. \tag{5}$$

For the 2-species case, assuming a Brownian motion model of evolution, $\boldsymbol{\Sigma}$ is just $\begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}$. Thus, substituting values for $\mathbf{X}, \mathbf{y}$, and $\boldsymbol{\Sigma}$ into Equation (4), we have

$$\hat{\boldsymbol{\beta}}_{\mathbf{GLS}} = \left(\begin{pmatrix} 1 & x_a \\ 1 & x_b \end{pmatrix}' \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & x_a \\ 1 & x_b \end{pmatrix} \right)^{-1}$$

$$\times \begin{pmatrix} 1 & x_a \\ 1 & x_b \end{pmatrix}' \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

$$= \left(\begin{pmatrix} \frac{1}{v_1} & \frac{1}{v_2} \\ \frac{x_a}{v_1} & \frac{x_b}{v_2} \end{pmatrix} \begin{pmatrix} 1 & x_a \\ 1 & x_b \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 \\ x_a & x_b \end{pmatrix} \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{v_1} + \frac{1}{v_2} & \frac{x_a}{v_1} + \frac{x_b}{v_2} \\ \frac{x_a}{v_1} + \frac{x_b}{v_2} & \frac{x_a^2}{v_1} + \frac{x_b^2}{v_2} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ x_a & x_b \end{pmatrix} \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

$$= \frac{1}{(x_b - x_a)^2} \begin{pmatrix} v_1 x_b^2 + v_2 x_a^2 & -v_1 x_b - v_2 x_a \\ -v_1 x_b - v_2 x_a & v_2 + v_1 \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & 1 \\ x_a & x_b \end{pmatrix} \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

$$= \frac{1}{x_b - x_a} \begin{pmatrix} v_1 x_b & -v_2 x_a \\ -v_1 & v_2 \end{pmatrix} \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix}$$

$$= \frac{1}{x_b - x_a} \begin{pmatrix} x_b & -x_a \\ -1 & 1 \end{pmatrix} \begin{pmatrix} y_a \\ y_b \end{pmatrix} \begin{pmatrix} \frac{x_b y_a - x_a y_b}{x_b - x_a} \\ \frac{y_b - y_a}{x_b - x_a} \end{pmatrix}.$$

We have now arrived at the important conclusion: The estimator of the regression slope for GLS is the same as that for PIC, arrived at previously. This is seen by comparing the second element of the GLS result (the slope element of vector $\boldsymbol{\beta}_{\mathbf{GLS}}$) above, with the PIC estimator $\boldsymbol{\beta}_{\mathbf{PIC}}$ in equation (2). Comparing the slope parameters for the two-taxon case does not mean that the 2 estimators are always identical for other sets of trees and trait data. This is the result that is proved in Appendix 1, Theorem A1. Further, the GLS intercept can also be recovered from the PIC analysis (Garland et al. 1993; Garland and Ives 2000), further confirming that the 2 modelling approaches are equivalent. This result is confirmed in Appendix 1 in Theorem A3.

ADVANTAGES AND LIMITATIONS

Algebraic evaluation of the equivalence of PIC and GLS for more than 2 species becomes tedious. For example, for the 3-species case with species a, b, and c and

with a and b more closely related to each other than to c, $\hat{\beta}_{\mathbf{PIC}}$ has 16 terms in the numerator and 12 terms in the denominator. $\hat{\beta}_{\mathbf{GLS}}$ has 21 terms in the numerator and 15 terms in the denominator. Few terms can be cancelled out, although the equivalence can easily be checked by substitution.

Compared with the calculation of $\hat{\beta}_{\mathbf{PIC}}$, computation of $\hat{\beta}_{\mathbf{GLS}}$ (above) is more tiresome and involves the inversion of Σ. However, the result is equivalent and comes with the added bonus of the estimator for the intercept. It is convenient that the PIC algorithm does not require an explicit inversion of the phylogenetic variance–covariance matrix Σ (i.e., this inversion is implicit in the algorithm rather than being calculated directly). In any case, inverting the phylogenetic variance–covariance matrix is generally a stable computation, as phylogenetic variance–covariance matrices are not singular. PIC may be particularly attractive when dealing with very large data sets as the inversion of large matrices can be computationally difficult. However, that advantage may be somewhat offset by the initial computation of the contrasts.

DISCUSSION

Because of the equivalence relationship between GLS and PIC, we are now in a better position to understand the behavior of PIC. The first term in Equation (3) represents the systematic part of the model for y in terms of the parameters of β and the explanatory variable X. The second term, ε, is the term of interest when trying to understand the role of phylogenetic covariance due to common ancestry. As stated above, ε is assumed to be Gaussian with zero mean and covariance $\sigma^2\Sigma$, where σ^2 is a scaling constant and Σ is the phylogenetic variance–covariance matrix. So $\sigma^2\Sigma$ describes the variance–covariance structure of the residuals. Σ and σ^2 enter into the calculation of $\widehat{\mathbf{Var}}(\hat{\beta})$, which is the estimate of the variance of $\hat{\beta}$. The true estimator of the variance of $\hat{\beta}$, $\mathbf{Var}(\hat{\beta})$ has the following form (Rao and Toutenburg 1995):

$$\mathbf{Var}(\hat{\beta}) = \sigma^2(X\Sigma^{-1}X)^{-1}, \tag{6}$$

where σ^2 has the unbiased estimator

$$s^2 = \frac{(y - X\hat{\beta})'\Sigma^{-1}(y - X\hat{\beta})}{n - k}, \tag{7}$$

whereit n is the sample size, that is, the number of species and k is the column rank of X. Hence, the estimator for the variance of $\hat{\beta}$ is just:

$$\widehat{\mathbf{Var}}(\hat{\beta}) = s^2(X\Sigma^{-1}X)^{-1}. \tag{8}$$

If the magnitude of the slope term in β is large relative to σ^2 and $\widehat{\mathbf{Var}}(\hat{\beta})$, then the existence of phylogenetic covariance will have little effect on the estimation and inference for β. Alternatively, if the magnitude of the β slope is small and σ^2 is moderate to large, estimation and inference for β will be strongly affected by phylogenetic covariance. This means that if the strength of the effect of X on y is large relative to the residual error, it does not much matter whether we estimate β by including phylogenetic covariance (as in GLS) or by ignoring it (as in OLS). In this case, conclusions using GLS or OLS will be very similar. It is important to understand that hypotheses about the relationship between X and y are concerned with the systematic part of the regression model and that this is distinct from hypotheses about phylogenetic effects arising from common ancestry, which are concerned with covariance of the residuals. Failure to understand this point may have caused some of the confusion in the literature (e.g., Harvey et al. 1995a, 1995b; Westoby et al. 1995a, 1995b, 1995c).

Another issue raised by the correspondence of GLS and PIC analyses is the interpretation of independent contrasts calculated for explanatory variables. The GLS model has the usual regression assumption that the explanatory variables are fixed and without error. The response y is a random variable conditioned on the explanatory variables. Hence, the variance–covariance matrix Σ applies to the response variable only. Application of the PIC algorithm to explanatory variables should be regarded as a mathematical idiosyncrasy in order to obtain $\hat{\beta}_{\mathbf{PIC}}$. Note that this interpretation is at odds with Felsenstein's (1985) derivation. In that paper, 2 traits were assumed to have evolved along a common phylogeny, so PICs were calculated for each trait and then OLS regression analysis was conducted (through the origin). However, even in such a case, the OLS regression assumptions imply that the contrasts for the explanatory variable are assumed fixed and without error. Of course, many (perhaps most) uses of regression in comparative biology are on traits that have evolved and exhibit phylogenetic signal (or covariance) and may be measured with some error. Such uses of OLS or GLS regression in these cases is generally permitted, although the magnitude of the measurement error in explanatory variable should be small compared with the magnitude of error in the response variable. If it is known that there is serious error in measurement of explanatory variables, then an "errors-in-variables" model should be used so that phylogenetic covariance is incorporated into both explanatory and response variables (see, e.g., Carroll et al. 2006; Ives et al. 2007). The situation is different for correlation analyses: Hypothesis tests involving Pearson correlation coefficients assume that the data are bivariate Normal, so both the explanatory and response variables are considered as random variables. Construction of PICs is then appropriate for the (noncentral) Pearson correlation analysis of 2 trait variables under Felsenstein's (1985) evolutionary model.

It has been suggested that different branch lengths can be used for different characters in a PIC multiple regression (Garland et al. 1992). By the Gauss–Markov–Aitken theorem, the GLS estimator for β (and by implication the PIC estimator for β since they are identical) is the

minimum variance unbiased estimator (MVUE). That is, of all the unbiased estimates of β, the GLS estimator has the minimum variance. Since the estimator is linear, it is also the best linear unbiased estimator (BLUE) when Σ is known (e.g., Mittelhammer et al. 2000, p. 43), where "best" means the estimator has the lowest mean square error of all linear estimators (Note also that the OLS estimator, β_{OLS}, is also unbiased but does not have minimum variance when $\Sigma \neq I$, the identity matrix (see Rohlf 2006).). Failing to use the same branch lengths for the explanatory variables implies that the resulting estimator is no longer MVUE and BLUE and is not recommended. Note also that the MVUE and BLUE properties do not require that the residuals are Gaussian. This assumption is only necessary if hypothesis tests are required. If the residuals are not Gaussian and hypothesis tests are required, then different methods are required (e.g., Ives and Garland 2010).

It is clear that we should do our best to use the MVUE estimator, which requires knowledge of Σ. The methods and results of phylogenetic systematics give us ways to estimate Σ, however, we must ask how uncertainty in Σ can affect the properties of the estimators. This topic has received little attention in the biological literature. The general term for the use of GLS methods in comparative biology is "phylogenetic" generalized least squares (PGLS, e.g., Rohlf 2001). However, there is nothing statistically special about the use of GLS methods for phylogenetic data per se. What is important about the use of GLS in comparative biology is that the phylogenetic variance–covariance matrix (Σ) is actually an estimate of the unobserved true variance–covariance matrix. Thus, we should restrict the use of Σ to refer to the true phylogenetic variance–covariance matrix and the estimate of this matrix is termed $\hat{\Sigma}$. We shall use this notation for the remainder of the paper. PGLS is identical to methods that econometricians term feasible generalized least squares or estimated generalized least squares (EGLS) (Mittelhammer et al. 2000).

The necessary conditions for EGLS to have the same properties as GLS (statistical consistency, unbiasedness, efficiency with respect to the OLS estimator, and minimum variance) have been discussed by Mittelhammer et al. (2000) and are summarized below. The EGLS estimator can be decomposed as in Equation (9):

$$\hat{\beta}_{EGLS} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$$
$$= \beta + (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\varepsilon, \qquad (9)$$

where β and ε are defined as for Equation (3). $\hat{\beta}_{EGLS}$ is an unbiased estimate of β if the last term in Equation (9) (i.e., the regression of residuals on X) is 0, and this occurs so long as the probability distribution of ε is symmetrical, so that the distributions of ε and $-\varepsilon$ are identical and centered at 0. This property is guaranteed by the Brownian motion assumption. However, if the underlying evolutionary model is a more complicated diffusion process with a drift component, asymmetry in ε and $-\varepsilon$ will occur and $\hat{\beta}_{EGLS}$ will be biased for β.

For the asymptotic properties of $\hat{\beta}_{EGLS}$ to be the same as those for $\hat{\beta}_{GLS}$, we require 2 additional conditions:

$$n^{-1}X'\hat{\Sigma}^{-1}X - n^{-1}X'\Sigma^{-1}X \xrightarrow{p} 0,$$
$$n^{-\frac{1}{2}}X'\hat{\Sigma}^{-1}\varepsilon - n^{-\frac{1}{2}}X'\Sigma^{-1}\varepsilon \xrightarrow{p} 0, \qquad (10)$$

where $\xrightarrow{p}$ describes convergence in probability (see Casella and Berger 2002). Equations (10) imply that $n^{-1}X'\hat{\Sigma}^{-1}X$ can replace $n^{-1}X'\Sigma^{-1}X$ in the limit, and the limiting distributions of $n^{-\frac{1}{2}}X'\hat{\Sigma}^{-1}\varepsilon$ and $n^{-\frac{1}{2}}X'\Sigma^{-1}\varepsilon$ are the same. Of particular importance is that consistent estimation of the elements of Σ with estimator $\hat{\Sigma}$ does not guarantee the properties described in Equation (10). This is because each element in matrices $n^{-1}X'\Sigma^{-1}X$ and $n^{-\frac{1}{2}}X'\Sigma^{-1}\varepsilon$ may depend on all the unique entries in Σ. Further, the entries in Σ are derived from the shared branch lengths among taxa so from a practical viewpoint, consistent estimation of all branch lengths is also a necessary (though not sufficient) condition. It should be noted that the conditions in Equation (10) cannot be checked with cross-species data; there is no sense of convergence of $\hat{\Sigma}$ to Σ as n, the number of species in the data set, increases. This is because each new species adds new covariance parameters to Σ. However, it is possible to consider convergence to Σ if we define convergence in terms of the addition of new sequence data to the phylogeny estimation data set rather than of adding new species to the trait data set.

Given the usual regularity conditions for GLS, and the extra conditions in Equations (10), it follows that $\hat{\beta}_{EGLS}$ is generally asymptotically more efficient than the OLS estimator $\hat{\beta}_{OLS}$. The proof of this relies on comparing the covariance matrices of the limiting distributions of the bias for $\hat{\beta}_{OLS}$ and $\hat{\beta}_{EGLS}$ (see Mittelhammer et al. 2000). This result depends on anticipating the true structure of Σ and the correctness of the convergence properties of $\hat{\Sigma}$ in Equations (10). If $\hat{\Sigma}$ is incorrectly designed, the finite sampling properties of $\hat{\beta}_{EGLS}$ can be "considerably worse" than those of $\hat{\beta}_{OLS}$ (Mittelhammer et al. 2000). In particular, if Σ has been misspecified (e.g., because the underlying tree topology or model of evolution is drastically wrong), $\hat{\beta}_{OLS}$ can be superior to $\hat{\beta}_{EGLS}$ (and therefore $\hat{\beta}_{PIC}$) even asymptotically. Fortunately, there has been some study of effects of misspecification of Σ. Revell (2010) demonstrates by computer simulation how $\hat{\beta}_{GLS}$ can be inferior to $\hat{\beta}_{OLS}$ in the situation where the true covariance structure is independence, and the assumed covariance matrix is derived from a phylogeny. Rao and Toutenburg (1995) provide analytical expressions for the bias and loss of efficiency introduced into $\hat{\beta}$ and $\hat{\sigma}^2$ by misspecification of $\hat{\Sigma}$, concentrating on the particular case where the true $\Sigma = I$ (the Identity matrix, meaning values of 1 on the diagonal for within-species variance, and 0 for all off-diagonal covariances). Inaccuracies in the topology have the most effect when species are moved across

the root. Changes closer to the tips of the tree are less drastic in their effects (see Martins and Housworth 2002; Symonds 2002). If branch lengths are transformed so that there is no relationship between the magnitude of contrasts and their standard deviation, Type I error rates are acceptable (Diaz-Uriarte and Garland 1996). In studies of real species and traits, we usually do not know how accurate our estimation of Σ may be. However, we can assess the robustness of our conclusions by running several analyses that incorporate alternative phylogenetic hypotheses (Donoghue and Ackerly 1996).

One recommendation that can be made is that both $\hat{\boldsymbol{\beta}}_{\mathbf{OLS}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{PIC}}$ (or $\hat{\boldsymbol{\beta}}_{\mathbf{GLS}}$) should be reported in studies where the phylogenetic relationships of the study species are only poorly known or if the phylogeny is suspected to be in severe error. This is contrary to the advice given by Freckleton (2009). Although we agree with this author in general on this issue, we believe that the poor frequentist properties of EGLS when $\hat{\boldsymbol{\Sigma}}$ is misspecified suggest that reporting the OLS estimates is useful in such problem cases. If the inferences for the OLS and GLS models differ in their conclusions, then not much can be said, except that a better estimate of Σ may be required. An alternative, suggested by a reviewer, is to adopt a more flexible correlation structure for the data such as the single parameter transformations using Grafen's ρ (Grafen 1989), Pagel's λ (Pagel 1999) or the Ornstein–Uhlenbeck model (Martins and Hansen 1997; Blomberg et al. 2003). Models can then be fitted using maximum likelihood (ML) or restricted maximum likelihood (REML) methods. Although such single parameter transformations can be very useful in controlling some types of misspecification of branch lengths (e.g., Revell 2010), they cannot account for errors in tree topology. Misspecification of the tree topology is still problematic for comparative studies.

It has been recently pointed out that many models in phylogenetic comparative analysis can be subsumed within the framework of ordinary linear mixed-effects models or generalized linear mixed-effects models (Hadfield and Nakagawa 2010) and that many models can be fitted efficiently with currently available software using REML (Gilmour et al. 2009; Pinheiro et al. 2009) or Bayesian methods (Hadfield 2010). It has further been argued that REML is superior to GLS because of the perceived bias and inflexibility of GLS (Hadfield and Nakagawa 2010). However, in the situation where there are no random effects, such as in Equation (3), and the variance–covariance matrix $\hat{\boldsymbol{\Sigma}}$ is fixed (i.e., the matrix does not contain parameters to be estimated, such as Grafen's ρ or Pagel's λ), the REML estimator is the same as the GLS estimator. Indeed, the key innovation of GLS for dealing with correlated data is the mathematical foundation for all modern methods for modeling multivariate Normal data. As an example, the function gls in the nlme package for R uses ML or REML to fit models like Equation (3), but the underlying code still uses the GLS "trick" of premultiplying the data matrix by $\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}}$ and applying Equation (5) (Pinheiro and Bates

2000). In practice, sparse matrix methods are used to perform this trick (Henderson 1976; Quaas 1976; Bates and DebRoy 2004; Hadfield and Nakagawa 2010) to avoid directly inverting $\hat{\boldsymbol{\Sigma}}$, which can be difficult when sample sizes are large.

A natural way of dealing with phylogenetic uncertainty, which avoids some of the problems of frequentist criteria for evaluating GLS estimators, is to use Bayesian methods (e.g., Box and Tiao 1973; Bernardo and Smith 2000; Carlin and Louis 2008). In the Bayesian paradigm, all variables are considered random, and variables each have a prior distribution that represents a statement of the researcher's prior knowledge and uncertainty (Gelman et al. 2003). For Bayesian regression models, Gaussian priors with zero mean and a large variance can be used for $\boldsymbol{\beta}$ parameters. A uniform prior can be used for σ (Gelman 2006).

Furthermore, whether taking a frequentist or a Bayesian approach, we can use a Bayes estimator of $\boldsymbol{\beta}$. Bayes estimators are estimators that minimize the posterior expected loss (Bernardo and Smith 2000). For simple loss functions, such as quadratic, linear, or 0–1 loss functions, the Bayes estimators are the posterior mean, median, and mode, respectively. Bayes estimators have the advantage that they very often have excellent frequentist properties (Robert 2007), so even if researchers do not wish to formally adopt the Bayesian paradigm, Bayes estimators can still be very useful. In a phylogenetic context, we envisage a collection of possible trees, perhaps resulting from a phylogenetic analysis using MrBayes (Huelsenbeck and Ronquist 2001) or BEAST (Drummond and Rambaut 2007). This distribution of trees can be sampled using the Metropolis–Hastings algorithm, allowing us to integrate over all trees in the collection, weighted in accordance to their posterior probability of generating the comparative data. Huelsenbeck and Rannala (2003) describe a similar method for estimating character correlations and the phylogeny simultaneously, however, our approach relies on the a priori existence of a collection of trees. This approach can work well and is easily implemented in available software such as BUGS (Lunn et al. 2000) or JAGS (Plummer 2003) (See Appendix 2). The Bayes estimators for $\boldsymbol{\beta}$ can be calculated directly from the posterior distribution of $\boldsymbol{\beta}$, with associated credible intervals. We have also found that various elaborations of the basic regression model are possible (de Villemereuil and Blomberg, unpublished manuscript).

REFERENCES

Bartlett M.S. 1951. An inverse matrix adjustment arising in discriminant analysis. Ann. Math. Stat. 22:107–111.

Bates D.M., DebRoy S. 2004. Linear mixed models and penalized least squares. J. Multivariate Anal. 91:1–17.

Bernardo J.M., Smith A.F. 2000. Bayesian theory. New York: Wiley.

Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution. 57:717–745.

Box G.E.P., Tiao G.C. 1973. Bayesian inference in statistical analysis. Reading (MA): Addison-Wesley.

Carlin B.P., Louis T.A. 2008. Bayesian methods for data analysis. 3rd ed. Boca Raton (FL): CRC Press. p. 552.

Carroll R.J., Ruppert D., Stefanski L.A., Crainiceanu C.M. 2006. Measurement error in nonlinear models. A modern perspective. (Monographs on statistics and applied probability; 2nd ed.). Boca Raton (FL): Chapman & Hall.

Casella G., Berger R.L. 2002. Statistical inference. Advanced Series. 2nd ed. Pacific Grove (CA): Duxbury.

Diaz-Uriarte R., Garland T.J. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Syst. Biol. 45:27–47.

Donoghue M.J., Ackerly D.D. 1996. Phylogenetic uncertainties and sensitivity analyses in comparative biology. Philos. Trans. Biol. Sci. 351:1241–1249.

Doughty P. 1996. Statistical analysis of natural experiments in evolutionary biology: comments on recent criticisms of the use of comparative methods to study adaptation. Am. Nat. 148:943–956.

Drummond A., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Freckleton R.P. 2009. The seven deadly sins of comparative analysis. J. Evol. Biol. 22:1367–1375.

Freckleton R.P., Harvey P.H., Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of the evidence. Am. Nat. 160:712–726.

Garland T., Ives A.R. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Garland T. Jr., Midford P.E., Ives A.R. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am. Zool. 39:39:374–388.

Garland T.G., Harvey P.H., Ives A.R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

Garland T.J., Dickerman A.W., Janis C.M., Jones J.A. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models (Comment on an article by Browne and Draper). Bayesian Anal. 1:515–533.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. 2003. Bayesian data analysis. Boca Raton (FA): Chapman & Hall.

Gilmour A., Gogel B., Cullis B.R., Thompson R. 2009. ASReml User Guide Release 3.0. Hemel Hempstead (UK): NSW Agriculture.

Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. London Ser. B Biol. Sci. 326:119–157.

Hadfield J. 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. J. Stat. Softw. 33: 1–22.

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.

Harvey P.H., Pagel. M.D. 1991. The comparative method in evolutionary biology. Oxford Series in Ecology and Evolution. Oxford: Oxford University Press.

Harvey P.H., Read A.F., Nee S. 1995a. Further remarks on the role of phylogeny in comparative ecology. J. Ecol. 83:733–734.

Harvey P.H., Read A.F., Nee S. 1995b. Why ecologists need to be phylogenetically challenged. J. Ecol. 83:535–536.

Henderson C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics. 32:69–83.

Horn R.A., Johnson C.R. 1985. Matrix analysis. Cambridge (UK): Cambridge University Press.

Huelsenbeck J.P., Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution. 57:1237–1247.

Huelsenbeck J.P., Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Ives A.R., Garland T. Jr. 2010. Phylogenetic logistic regression for binary dependent variables. Syst. Biol. 59:9–26.

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.

Ives A.R., Zhu J. 2006. Statistics for correlated data: phylogenies, space, and time. Ecol. Appl. 16:20–32.

Jenks R.D., Sutor R.S. 1992. Axiom—the scientific computation system. Berlin: Springer.

Kluge A.G. 2005. Testing lineage and comparative methods for inferring adaptation. Zool. Scr. 34:653–663.

Leroi A.M., Rose M.R., Lauder G.V. 1994. What does the comparative method reveal about adaptation? Am. Nat. 143:381–402.

Lunn D., Spiegelhalter D., Thomas A., Best N. 2009. The BUGS project: evolution, critique and future directions (with discussion). Stat. Med. 28:3049–3082.

Lunn D.J., Spiegelhalter D., Thomas A., Best N. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. 10:325–327.

Martins E.P. 2000. Adaptation and the comparative method. Trends Ecol. Evol. 15:296–299.

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.

Martins E.P., Housworth. E.A. 2002. Phylogeny shape and the phylogenetic comparative method. Syst. Biol. 51:873–880.

Mittelhammer R.C., Judge G.C., Miller D.J. 2000. Econometric foundations. Cambridge (UK): Cambridge University Press.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Pinheiro J., Bates D., DebRoy S., Sarkar D., the R Core team. 2009. nlme: linear and nonlinear mixed effects models (R package). Version 3.1-91.

Pinheiro J. C. and Bates D. M. 2000. Mixed-effects models in S and S-PLUS. Statistics and Computing. New York: Springer.

Plummer M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik K., Leisch F., Zeileis A., editors. Proceedings of the Third International Workshop on Distributed Statistical Computing (DSC 2003), Vienna (Austria). p. 1–10.

Quaas R. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics. 32:949–953.

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/. ISBN 3-900051-07-0.

Rao C.R., Toutenburg H. 1995. Linear models. Least squares and alternatives. Berlin: Springer Series in Statistics Springer.

Revell L.J. 2010. Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. 1:319–329.

Robert C.P. 2007. The Bayesian choice. Springer Texts in Statistics. 2nd ed. New York: Springer.

Rohlf F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution. 55:2143–2160.

Rohlf F.J. 2006. A comment on phylogenetic correction. Evolution. 60:1509–1515.

Seroul R. 2000. Programming for mathematicians. Universitext. Berlin: Springer.

Symonds M.R.E. 2002. The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. Syst. Biol. 51:541–553.

Westoby M., Leishman M., Lord J. 1995a. Further remarks on phyloge-
    netic correction. J. Ecol. 83:727–729.
Westoby M., Leishman M., Lord J. 1995b. Issues of interpretation after
    relating comparative datasets to phylogeny. J. Ecol. 83:892–893.
Westoby M., Leishman M.R., Lord J.M. 1995c. On misinterpreting the
    'phylogenetic correction'. J. Ecol. 83:531–534.

APPENDICES

*Appendix 1: Proof of the Equivalence of the GLS and PIC
Estimators*

For a given phylogenetic tree T with $n \geqslant 1$ species,
let Σ denote the corresponding covariance matrix of
the species, which we assume is invertible. The invert-
ibility assumption holds for all positive-definite ma-
trices (Horn and Johnson 1985) of which phylogenetic
variance–covariance matrices are a subset. From a prac-
tical viewpoint, some positive-definite matrices are dif-
ficult to invert using numerical methods, such as from
trees that are very large, or have some very short inter-
nal branches. However, they are invertible in theory. x
and y are column n-vectors of explanatory and response
variables, respectively. Note that Σ and Σ^{-1} are both
symmetric. Let 1 denote a column n-vector of ones and
define the following scalars:

$$S = 1'\Sigma^{-1}1,$$
$$A = 1'\Sigma^{-1}x = x'\Sigma^{-1}1,$$
$$B = x'\Sigma^{-1}x,$$
$$C = 1'\Sigma^{-1}y = y'\Sigma^{-1}1,$$
$$D = x'\Sigma^{-1}y.$$

Note that S is the sum of all entries in Σ^{-1}. The PIC esti-
mator for the slope of the regression is given by

$$\hat{\beta}_{\text{PIC}} = \frac{X'_{\text{PIC}}y_{\text{PIC}}}{X'_{\text{PIC}}X_{\text{PIC}}},$$

where X_{PIC} and y_{PIC} are column vectors (length
$n - 1$) of PICs, calculated for the explanatory and re-
sponse variables, respectively.

The GLS estimators for the intercept and slope of the
regression are given by

$$\hat{\beta}_{\text{GLS}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$
$$= \begin{pmatrix} S & A \\ A & B \end{pmatrix}^{-1} \begin{pmatrix} C \\ D \end{pmatrix}$$
$$= \frac{1}{SB - A^2} \begin{pmatrix} BC - AD \\ SD - AC \end{pmatrix}. \qquad (A.1)$$

Since we are only interested in the estimated slope, let
$\hat{\beta}_s = \frac{SD - AC}{SB - A^2}$.

Theorem 1. *The PIC and GLS estimators of the regression
slope are identical. That is, $\hat{\beta}_{\text{PIC}} = \hat{\beta}_s$.*

Proof.   We will prove that

$$X'_{\text{PIC}}y_{\text{PIC}} = D - AC/S. \qquad (A.2)$$

Substituting X_{PIC} for y_{PIC} in Equation (A.2) gives
$X'_{\text{PIC}}X_{\text{PIC}} = B - A^2/S$, and the result will then follow.
In order to prove Equation (A.2), we also need to show
that the weighted means obtained from the tree satisfy

$$X_{\text{m}} = A/S \text{ and} \qquad (A.3)$$
$$y_{\text{m}} = C/S, \qquad (A.4)$$

with associated variance

$$V_{\text{m}} = 1/S. \qquad (A.5)$$

$\square$

Example 2   For a phylogenetic tree with just 2 species,
then

$$\Sigma = \begin{pmatrix} v_1 & 0 \\ 0 & v_2 \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{pmatrix},$$

$$S = \frac{1}{v_1} + \frac{1}{v_2}, A = \frac{x_1}{v_1} + \frac{x_2}{v_2}, B = \frac{x_1^2}{v_1} + \frac{x_2^2}{v_2},$$

$$C = \frac{y_1}{v_1} + \frac{y_2}{v_2}, D = \frac{x_1 y_1}{v_1} + \frac{x_2 y_2}{v_2},$$

$$X_{\text{m}} = \frac{x_1 v_2 + x_2 v_1}{v_1 + v_2}, y_{\text{m}} = \frac{y_1 v_2 + y_2 v_1}{v_1 + v_2}, V_{\text{m}} = \frac{v_1 v_2}{v_1 + v_2},$$

$$X_{\text{PIC}} = \left(\frac{x_1 - x_2}{\sqrt{v_1 + v_2}} \right) \quad \text{and} \quad y_{\text{PIC}} = \left(\frac{y_1 - y_2}{\sqrt{v_1 + v_2}} \right).$$

Therefore,

$$SD - AC = \frac{x_1 y_1}{v_1^2} + \frac{x_2 y_2}{v_2^2} + \frac{x_1 y_1 + x_2 y_2}{v_1 v_2} - \frac{x_1 y_1}{v_1^2} - \frac{x_2 y_2}{v_2^2}$$
$$- \frac{x_1 y_2 + x_2 y_1}{v_1 v_2}$$
$$= \frac{(x_1 - x_2)(y_1 - y_2)}{v_1 v_2},$$

which implies that

$$D - \frac{AC}{S} = \frac{(x_1 - x_2)(y_1 - y_2)}{v_1 + v_2}.$$

Thus, Equations (A.2)–(A.5) hold for a tree with 2 nodes.
Thus, we have proved again the example in the intro-
duction.

The proof is by strong induction on the size of the
phylogenetic tree. Every phylogenetic tree with at least 2
species is represented as a combination of 2 smaller phy-
logenetic trees; we assume that Equations (A.2)–(A.5)
hold for the smaller trees and prove that they must then
also hold for the complete tree.

Before considering the details of the inductive step, we consider phylogenetic trees with only one species (the basis case). Considered as an isolated node, such a tree has a noninvertible covariance matrix. Thus, to avoid division by zero, we assume that the tree contains a branch of weight $v > 0$ from some unspecified ancestor. This is a departure from the convention that every phylogenetic tree has at least two distinct branches from the base node. The covariance matrix is then simply (v), with inverse $\left(\frac{1}{v}\right)$, and Equations (A.2)–(A.5) hold trivially. Note that these equations are well defined although $\hat{\beta}_{PIC}$ and $\hat{\beta}_s$ are not when $n = 1$.

For some $n \geqslant 2$, we now suppose that every phylogenetic tree with fewer than n species has an invertible covariance matrix and satisfies (A.2)–(A.5). We consider an arbitrary tree T containing n species and show that T has an invertible covariance matrix and satisfies (A.2)–(A.5). Since T (and every other tree with more than 1 species) is assumed to have at least 2 distinct branches from its base node, we may regard T as the combination of 2 distinct phylogenetic trees T_1 and T_2, each with fewer than n species, connected to the base node of T by branches of weight d_1 and d_2, respectively. Note that $d_i = 0$ is allowed, implying that the base node of T_i is identical to the base node of T, so there is no upper limit on the number of branches leading from the base node of T. If T_i contains a single species, for $i \in \{1, 2\}$, then the branch from the base node of T to that species is regarded as part of T_i; hence $d_i = 0$.

For $i = 1, 2$, define n_i, $\mathbf{\Sigma}_i$, A_i, B_i, C_i, D_i, $\mathbf{x}_i$, $\mathbf{y}_i$, and $\mathbf{1}_i$ analogously to n, $\mathbf{\Sigma}$, A, B, C, D, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{1}$; note that $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ are invertible by the inductive assumption. Thus, we have

$$n = n_1 + n_2, \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \text{ and } \mathbf{1} = \begin{pmatrix} \mathbf{1}_1 \\ \mathbf{1}_2 \end{pmatrix}.$$

By our inductive assumption, Equations (A.2)–(A.5) hold for both T_1 and T_2.

Let $J = \mathbf{11}'$ be an $n \times n$ matrix of ones (with J_1 and J_2 defined analogously). By the Sherman–Morrison formula (Bartlett 1951), we have that

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_1} + d_1 J_1 & 0 \\ 0 & \mathbf{\Sigma_2} + d_2 J_2 \end{pmatrix}, \text{ and hence}$$

$$\mathbf{\Sigma^{-1}} = \begin{pmatrix} \mathbf{\Sigma_1^{-1}} - \frac{d_1 \mathbf{\Sigma_1^{-1}} J_1 \mathbf{\Sigma_1^{-1}}}{1 + d_1 S_1} & 0 \\ 0 & \mathbf{\Sigma_2^{-1}} - \frac{d_2 \mathbf{\Sigma_2^{-1}} J_2 \mathbf{\Sigma_2^{-1}}}{1 + d_2 S_2} \end{pmatrix}.$$

This is well defined, so we have established that Σ is invertible.

Let $\mathbf{z} \in \{\mathbf{x}, \mathbf{y}, \mathbf{1}\}$. Then

$$\mathbf{1}'\mathbf{\Sigma^{-1}z} = \mathbf{1}_1'\mathbf{\Sigma_1^{-1}z_1} + \mathbf{1}_2'\mathbf{\Sigma_2^{-1}z_2} - \frac{d_1 \mathbf{1}_1'\mathbf{\Sigma_1^{-1}}\mathbf{1}_1 \mathbf{1}_1'\mathbf{\Sigma_1^{-1}z_1}}{1 + d_1 S_1}$$

$$- \frac{d_2 \mathbf{1}_2'\mathbf{\Sigma_2^{-1}}\mathbf{1}_2 \mathbf{1}_2'\mathbf{\Sigma_2^{-1}z_2}}{1 + d_2 S_2}$$

$$= \frac{\mathbf{1}_1 \mathbf{\Sigma_1^{-1}z_1}}{1 + d_1 S_1} + \frac{\mathbf{1}_2 \mathbf{\Sigma_2^{-1}z_2}}{1 + d_2 S_2}.$$

Thus,

$$S = \frac{S_1}{1 + d_1 S_1} + \frac{S_2}{1 + d_2 S_2},$$

$$A = \frac{A_1}{1 + d_1 S_1} + \frac{A_2}{1 + d_2 S_2}, \text{ and}$$

$$C = \frac{C_1}{1 + d_1 S_1} + \frac{C_2}{1 + d_2 S_2}.$$

Likewise,

$$\mathbf{x}'\mathbf{\Sigma^{-1}z} = \mathbf{x}_1'\mathbf{\Sigma_1^{-1}z_1} + \mathbf{x}_2'\mathbf{\Sigma_2^{-1}z_2} - \frac{d_1 \mathbf{x}_1'\mathbf{\Sigma_1^{-1}}\mathbf{1}_1 \mathbf{1}_1'\mathbf{\Sigma_1^{-1}z_1}}{1 + d_1 S_1}$$

$$- \frac{d_2 \mathbf{x}_2'\mathbf{\Sigma_2^{-1}}\mathbf{1}_2 \mathbf{1}_2'\mathbf{\Sigma_2^{-1}z_2}}{1 + d_2 S_2}.$$

Hence,

$$B = B_1 + B_2 - \frac{d_1 A_1^2}{1 + d_1 S_1} - \frac{d_2 A_2^2}{1 + d_2 S_2} \text{ and}$$

$$D = D_1 + D_2 - \frac{d_1 A_1 C_1}{1 + d_1 S_1} - \frac{d_2 A_2 C_2}{1 + d_2 S_2}.$$

By definition,

$$X_m = \frac{X_{m_1}(d_2 + V_{m_2}) + X_{m_2}(d_1 + V_{m_1})}{d_1 + d_2 + V_{m_1} + V_{m_2}},$$

and so by the inductive assumption,

$$X_m = \frac{\frac{A_1}{S_1}\left(d_2 + \frac{1}{S_2}\right) + \frac{A_2}{S_2}\left(d_1 + \frac{1}{S_1}\right)}{d_1 + d_2 + \frac{1}{S_1} + \frac{1}{S_2}}$$

$$= \frac{A_1(d_2 S_2 + 1) + A_2(d_1 S_1 + 1)}{S_1(d_2 S_2 + 1) + S_2(d_1 S_1 + 1)}$$

$$= \left(\frac{A_1}{d_1 S_1 + 1} + \frac{A_2}{d_2 S_2 + 1}\right) \Big/ \left(\frac{S_1}{d_1 S_1 + 1} + \frac{S_2}{d_2 S_2 + 1}\right)$$

$$= \frac{A}{S}.$$

Thus, Equation (A.3) holds; similarly for Equation (A.4).

Also by definition,

$$V_m = \frac{(d_2 + V_{m_2})(d_1 + V_{m_1})}{d_1 + d_2 + V_{m_1} + V_{m_2}}$$

$$= \frac{(d_2 + 1/S_2)(d_1 + 1/S_1)}{d_2 + 1/S_2 + d_1 + 1/S_1}$$

$$= \frac{(d_2 S_2 + 1)(d_1 S_1 + 1)}{S_1(d_2 S_2 + 1) + S_2(d_1 S_1 + 1)}$$

$$= 1 \Big/ \left(\frac{S_1}{d_1 S_1 + 1} + \frac{S_2}{d_2 S_2 + 1}\right)$$

$$= 1/S,$$

so Equation (A.5) holds.

We now just need to prove the inductive step for Equation (A.2).

From the PIC definition, we have

$$X_{PIC} = \begin{pmatrix} X_{PIC_1} \\ X_{PIC_2} \\ \frac{X_{m_1} - X_{m_2}}{\sqrt{V_{m_1} + V_{m_2} + d_1 + d_2}} \end{pmatrix} \text{ and}$$

$$y_{PIC} = \begin{pmatrix} y_{PIC_1} \\ y_{PIC_2} \\ \frac{y_{m_1} - y_{m_2}}{\sqrt{V_{m_1} + V_{m_2} + d_1 + d_2}} \end{pmatrix}.$$

Therefore,

$$X'_{PIC}y_{PIC} = X'_{PIC_1}y_{PIC_1} + X'_{PIC_2}y_{PIC_2}$$
$$+ \frac{(X_{m_1} - X_{m_2})(y_{m_1} - y_{m_2})}{V_{m_1} + V_{m_2} + d_1 + d_2}.$$

Since Equations (A.2)–(A.5) hold for T_1 and T_2, then

$$X'_{PIC}y_{PIC} = D_1 - \frac{A_1 C_1}{S_1} + D_2 - \frac{A_2 C_2}{S_2}$$
$$+ \frac{\left(\frac{A_1}{S_1} - \frac{A_2}{S_2}\right)\left(\frac{C_1}{S_1} - \frac{C_2}{S_2}\right)}{\frac{1}{S_1} + \frac{1}{S_2} + d_1 + d_2}$$
$$= D_1 + D_2 - \frac{d_1 A_1 C_1}{1 + d_1 S_1} - \frac{d_2 A_2 C_2}{1 + d_2 S_2}$$
$$- \frac{\left(\frac{A_1}{1 + d_1 S_1} + \frac{A_2}{1 + d_2 S_2}\right)\left(\frac{C_1}{1 + d_1 S_1} + \frac{C_2}{1 + d_2 S_2}\right)}{\left(\frac{S_1}{1 + d_1 S_1} + \frac{S_2}{1 + d_2 S_2}\right)}$$
$$= D - AC/S. \qquad \square$$

In addition to the slope estimate, we can use the weighted means X_m and y_m generated by the PIC algorithm to estimate the intercept.

Theorem 3. *The first element of $\hat{\boldsymbol{\beta}}_{GLS}$ is equivalent to $y_m - \hat{\beta}_{PIC}X_m$ (Garland et al. 1993).*

Proof. By Theorem 1 and Equations (A.3) and (A.4),

$$y_m - \hat{\beta}_{PIC}X_m = \frac{BC - AD}{SB - A^2}, \qquad (A.6)$$

where the right-hand side of Equation (A.6) corresponds to the first element of the vector in Equation (A.1). $\square$

Appendix 2: BUGS Code for Simple Bayesian Phylogenetic Regression

The following BUGS code shows how to implement Bayesian phylogenetic regression analyses for Open-BUGS (Lunn et al. 2009) or JAGS (Plummer 2003). It is assumed that there are N species, with explanatory variable X and response variable Y. It is assumed there are Ntree trees in the set of phylogenies derived from some Bayesian phylogenetics software (e.g., BEAST or MrBayes). These need to be converted to phylogenetic variance–covariance matrices for example in R Development Core Team (2011) using the vcv.phylo function in the ape package (Paradis et al. 2004). Also, since the matrices need to be inverted as part of the computation, it is more efficient to invert the matrices first, say in R. They should then be saved as part of the OpenBUGS data as a large 3-dimensional array of size Ntree $\times N \times N$. This array is referred to as invA in the code below. Anything after a "#" is a comment.

```
model {
# Linear regression and multivariate normal likelihood
for ( i in 1:N) {
    mu[i] <- beta0+beta1*X[i] # beta0 is the intercept
                              # and beta1 is the slope
                              # of the regression of
                              # Y on X
}
Y[1:N] ~ dmnorm (mu[] ,TAU[ ,]) # Y is multivariate normal
                                # with mean mu and precision
                                # matrix TAU
# Priors
beta0 ~ dnorm (0, 1.0E-06) # uninformative priors
beta1 ~ dnorm (0, 1.0E-06)
sigma ~ dunif (0, 10) # See Gelman (2006)
tau <- 1/sigma ^ 2 # tau is a scaling factor for the precision
                   # matrix
# We sample trees (precision matrices) with equal prior probability
# from invA
for (k in 1:Ntree) {
    p[k] <- 1/Ntree
}
```

```
# Tree sampling and variance-covariance matrix construction
K ~ dcat(p [])
for (i in 1:N) {
        for (j in 1:N) {
                TAU[i, j] <- tau*invA[i, j, K] # calculate the
                                               # precision matrix
                }
        }
}
```