

R^2 s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs

ANTHONY R. IVES*

Department of Integrative Biology, UW-Madison, Madison, WI 53706, USA

*Correspondence to be sent to: Department of Integrative Biology, UW-Madison, 430 Lincoln Dr., Madison, WI 53706, USA;
E-mail: arives@wisc.edu.

Received 14 August 2017; reviews returned 10 September 2018; accepted 12 September 2018
Associate Editor: Luke Harmon

Abstract.—Many researchers want to report an R^2 to measure the variance explained by a model. When the model includes correlation among data, such as phylogenetic models and mixed models, defining an R^2 faces two conceptual problems. (i) It is unclear how to measure the variance explained by predictor (independent) variables when the model contains covariances. (ii) Researchers may want the R^2 to include the variance explained by the covariances by asking questions such as “How much of the data is explained by phylogeny?” Here, I investigated three R^2 s for phylogenetic and mixed models. R^2_{resid} is an extension of the ordinary least-squares R^2 that weights residuals by variances and covariances estimated by the model; it is closely related to R^2_{gmm} presented by Nakagawa and Schielzeth (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4:133–142). R^2_{pred} is based on predicting each residual from the fitted model and computing the variance between observed and predicted values. R^2_{lik} is based on the likelihood of fitted models, and therefore, reflects the amount of information that the models contain. These three R^2 s are formulated as partial R^2 s, making it possible to compare the contributions of predictor variables and variance components (phylogenetic signal and random effects) to the fit of models. Because partial R^2 s compare a full model with a reduced model without components of the full model, they are distinct from marginal R^2 s that partition additive components of the variance. I assessed the properties of the R^2 s for phylogenetic models using simulations for continuous and binary response data (phylogenetic generalized least squares and phylogenetic logistic regression). Because the R^2 s are designed broadly for any model for correlated data, I also compared R^2 s for linear mixed models and generalized linear mixed models. R^2_{resid} , R^2_{pred} , and R^2_{lik} all have similar performance in describing the variance explained by different components of models. However, R^2_{pred} gives the most direct answer to the question of how much variance in the data is explained by a model. R^2_{resid} is most appropriate for comparing models fit to different data sets, because it does not depend on sample sizes. And R^2_{lik} is most appropriate to assess the importance of different components within the same model applied to the same data, because it is most closely associated with statistical significance tests. [Binomial regression; coefficient of determination; non-independent residuals; phylogenetic model; pseudo-likelihood.]

Researchers often want to calculate a coefficient of determination, an R^2 , to give a measure of the amount of variance in their data explained by a statistical model. For ordinary least-squares models (OLS), such as regression and analysis of variance (ANOVA), the R^2 is simple to calculate and interpret. Many types of models, however, assume that the errors among response variables are correlated. Phylogenetic generalized least squares models (PGLS) allow the possibility of phylogenetically related species being more similar to each other, leading to phylogenetic correlations in the errors (Felsenstein, 1985; Garland et al., 1992; Martins and Hansen, 1997). PGLS models are structurally similar to linear mixed models (LMMs) that include random effects to account for correlations in the residual variation; for example, LMMs can account for correlation between residuals of experimental replicates within the same block (Gelman and Hill, 2007; Bolker et al., 2009). Interpretation of an R^2 for models for discrete response variables, such as phylogenetic logistic regression models and generalized linear mixed models (GLMMs), adds a second complexity. Even a perfectly fitting model will have residual variation due to the discreteness of the data, which raises the question of how to formulate an R^2

when the variance in the data can never be completely explained.

Formally, correlated errors in statistical models for continuous or discrete data cause two issues for defining an R^2 . The first involves assessing the goodness-of-fit of predictor variables (fixed effects) in terms of the explained variance. For standard OLS models, the errors are assumed to be identical and independently distributed, and therefore, the variance in the residuals can be calculated directly to give the total variance that is not explained by the model. In models for correlated data, however, the errors are not independently distributed. Therefore, to calculate the “unexplained variance” given by the residuals, it is necessary to deal with the covariances among errors; applying the OLS R^2 to estimates from a model with covariances among errors gives values that are bounded below by $-\infty$ rather than zero (Judge et al., 1985, p. 32).

The second issue for defining an R^2 involves assessing the goodness-of-fit of the covariances (random effects) estimated in the model. For phylogenetic models, this is embodied by the question “How much of the data is explained by phylogeny?” The difficulty is that a phylogenetic model can be used to estimate the strength

of phylogenetic signal (the magnitude of the covariances) in the errors, but the phylogenetic signal does not directly quantify the proportion of the total (rather than residual) variance explained. For example, in PGLS Pagel's λ branch-length transform (Pagel, 1997; Housworth et al., 2004) partitions the variance of the residuals into phylogenetic and non-phylogenetic parts; this is similar to LMMs that partition variances into random effects and residuals. In contrast, in OLS the R^2 partitions the total variance in Y into that part explained by the predictor variables (fixed effects) and the residual variation. It is not immediately clear what it means in PGLS for a phylogeny to "explain" the total variance of the data in a way comparable to the predictor variables in OLS.

Here, I assess three R^2 s for models that specify non-zero covariances among errors. Although these definitions of R^2 s are broad enough to encompass any model specifying an error covariance matrix, I will focus on application to phylogenetic models for continuous (PGLS) and binary (PGLMM) data. In addition, I will compare the properties of the R^2 s applied to phylogenetic data to the properties of the R^2 s applied to mixed models. This comparison shows the generalizability of the R^2 s, and validates the R^2 s as viable measures of goodness-of-fit for a broad class of models. This is important, because R^2 s should make it possible to assess and compare as wide a range of models as possible (Kvalseth, 1985).

The three R^2 s are designed as partial R^2 s that compare a full model with a reduced model in which one or more of the parameters are removed; partial R^2 s measure the explained variance that is lost when the full model is reduced. The overall R^2 s are obtained by comparing the full model to the simplest reduced model in which there is only an intercept and the residuals are independent. Partial R^2 s have the advantage of making it possible to ask about the contribution of a single or subset of components to the fit of a model. This makes it possible to exclude coefficients in a model that are not of explicit interest; for example, many phylogenetic models for species traits include body size as one of the predictor variables to factor out body size, and partial R^2 s make it possible to assess the goodness-of-fit for the remaining predictor variables. By comparing a model with a phylogeny to a model without, partial R^2 s also make it possible to answer the question "How much of the data is explained by phylogeny?"

R^2 s can be assessed on multiple grounds (Kvalseth, 1985), and here I use three. First, does the R^2 give a good measure-of-fit of a model to data? To serve as a basis for assessment, I use the log-likelihood ratio (LLR) of the full and reduced models. The LLR approaches a χ^2 distribution for large samples and is therefore used for hypothesis tests of full versus reduce models (Judge et al., 1985). Also, the LLR is linearly related to the Akaike's Information Criterion (AIC) and other measures used for model selection (Burnham and Anderson, 2002).

Therefore, the LLR is a natural choice to assess R^2 s: a good R^2 should be monotonically related to the LLR. Second, can the R^2 identify the contribution of different components of the model, specifically that component attributed to predictor variables and that attributed to phylogenetic or random effects, to the overall model fit? The partial R^2 corresponding to the phylogenetic or random effect should give an indication of the statistical significance of these covariance structures in the data. Third, does the R^2 give similar values when applied to data generated by the same statistical process? If the values of R^2 when applied to data generated from the same statistical process are all similar, then the R^2 gives a precise measure of goodness-of-fit.

MATERIALS AND METHODS

Here, I present three R^2 s— R^2_{resid} , R^2_{prd} , and R^2_{lik} —that can be applied to a broad class of models in which the variance structure of the residuals contains covariances. As a strategy of attack, I will begin with a detailed discussion of R^2_{resid} , with the goal of explaining the challenges of defining an R^2 for correlated data as well as possible solutions. R^2_{resid} is based on the variance of residuals of a fitted model and is related to R^2_{glmm} (Nakagawa and Schielzeth, 2013), and contrasting R^2_{resid} with R^2_{glmm} generates a discussion of what partial R^2 s reveal about a fitted model. I will then give briefer presentations of R^2_{pred} and R^2_{lik} . All R^2 s are computed in the R package rr2 (Ives and Li, 2018).

$$R^2_{resid}$$

There is an extensive literature on R^2 s for GLMs and LMMs, and a growing literature for GLMMs (Buse, 1973; Cameron and Windmeijer, 1996, 1997; Kenward and Roger, 1997; Menard, 2000; Xu, 2003; Kramer, 2005; Edwards et al., 2008; Liu et al., 2008; Orellien and Edwards, 2008; Nakagawa and Schielzeth, 2013; Jaeger et al., 2017); this literature forms the basis for the R^2 s that can be applied to phylogenetic models. The three R^2 s take three different approaches to defining "explained variance", the same general approaches considered for LMMs by Xu (2003). The R^2 discussed first, R^2_{resid} (for residual variance), is similar to R^2_{glmm} presented by Nakagawa and Schielzeth (2013) and related work (Edwards et al., 2008; Jaeger et al., 2017; Nakagawa et al., 2017). Therefore, I will present R^2_{resid} first in application to GLMMs and then in application to phylogenetic models.

For the sake of exposition, consider the simple LMM for a species-specific response variable Y that depends on a predictor variable x and a random effect b that describes variance in the mean (intercept) response of

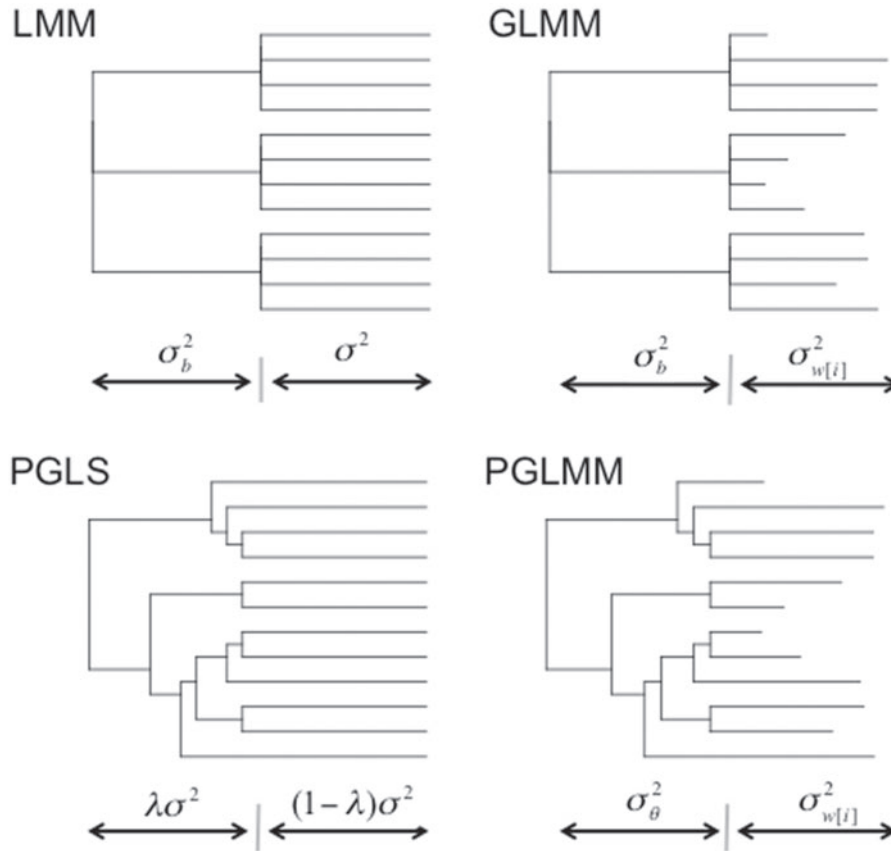


FIGURE 1. Depictions of the covariance matrices from LMM, PGLS, GLMM, and PGLMM models. In the covariance matrix for LMMs, $\sigma^2 \Sigma(\sigma_b^2) = \sigma_b^2 \Sigma_b + \sigma^2 \mathbf{I}$, the variance of the random effect σ_b^2 is scaled against the variance of the residual errors. In the PGLS with a Pagel's λ branch-length transform, the covariance matrix is $\sigma^2 \Sigma(\lambda) = \sigma^2 \lambda \Sigma_{BM} + \sigma^2 (1 - \lambda) \mathbf{I}$, in which λ determines the strength of phylogenetic signal in the residual errors. For GLMMs, the variance of the random effect is given by $\Sigma(\sigma_b^2) = \sigma_b^2 \Sigma_b$, and there is additional variance $\sigma_{w[i]}^2$ owing to the discreteness of the data. Similarly, for PGLMM, phylogenetic signal enters the model as a covariance matrix $\sigma_b^2 \Sigma_{BM}$, with additional variance $\sigma_{w[i]}^2$.

species from different taxonomic groups:

$$\begin{aligned}
 Y_i &= \beta_0 + \beta_1 x_i + b_{\text{taxon}[i]} + e_i \\
 b_{\text{taxon}[i]} &\sim \text{Gaussian}(\mathbf{0}, \sigma_b^2) \\
 e_i &\sim \text{Gaussian}(\mathbf{0}, \sigma_e^2)
 \end{aligned}
 \tag{1}$$

Here, assume there are n species ($i=1, \dots, n$) and m taxonomic groups, so that the random effect b has m levels representing m draws from a Gaussian distribution with mean 0 and variance σ_b^2 , and the function $\text{taxon}[i]$ maps species i into its appropriate taxonomic group. This model can be visualized as a phylogeny in which species within the same taxon have a common ancestor, and all common ancestors occur at the same height of the tree (Fig. 1, LMM). R_{glmm}^2 (Nakagawa and Schielzeth, 2013) is computed from the different sources of variance in Y_i . Specifically, for the model in equation (1) the conditional $R_{glmm(c)}^2$ is

$$R_{glmm(c)}^2 = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2}
 \tag{2}$$

given in terms of the estimates of σ_b^2 and σ_e^2 , and the estimate of the variance in x , $\hat{\sigma}_x^2 = \beta^2 \text{var}(x_i)$. This has the usual interpretation of R^2 s as one minus the unexplained residual variance relative to the total variance. $R_{glmm(c)}^2$ gives a measure of the total explained variance, where variance attributed to the random effects is considered as explained.

It is also useful to ask how much of the total explained variance can be attributed to random effects. Nakagawa and Schielzeth (2013) define the marginal R^2 corresponding to the fixed effects as

$$R_{glmm(m)}^2 = 1 - \frac{\hat{\sigma}_e^2 + \hat{\sigma}_b^2}{\hat{\sigma}_x^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2}
 \tag{3}$$

Here, the variance unexplained by the fixed effects now includes the variance of the random effect. Since we also want to know the R^2 corresponding to the random effects, we could extend these ideas to give a marginal $R_{glmm(r)}^2$ for the random effects:

$$R_{glmm(r)}^2 = 1 - \frac{\hat{\sigma}_e^2 + \hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2}
 \tag{4}$$

$R^2_{glmm(m)}$ and $R^2_{glmm(r)}$ give a partition of $R^2_{glmm(c)}$ into parts associated with the fixed and random effects, since $R^2_{glmm(m)} + R^2_{glmm(r)} = R^2_{glmm(c)}$. However, they are not partial R^2 s. Partial R^2 s depend on the difference between the full model fitted with all components, and a reduced model in which one or more components (e.g., the fixed or random effects) have been removed. For example, the partial R^2_{glmm} for the fixed effect x would be

$$R^2_{glmm.partial.x} = 1 - \frac{1 - R^2_{glmm}}{1 - R^2_{glmm.r}} = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_{e,r}^2} \left(\frac{\hat{\sigma}_{e,r}^2 + \hat{\sigma}_{e,r}^2}{\hat{\sigma}_x^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2} \right) \quad (5)$$

where subscripts containing “.r” correspond to the estimates of the variances from the reduced model in which x is excluded. The numerator and denominator of the term in parentheses are estimates of the total variance from the reduced and full models, respectively. Even though they estimate the same thing, their estimates differ, normally by only a little but potentially by a lot depending on the fits of the full and reduced models. Because this complicates comparisons between total and partial R^2 s, and between partial R^2 s for different reduced models, R^2_{resid} for GLMMs is defined using the sample variance $\text{var}(Y)$. Because $\text{var}(Y)$ is the same for full and all reduced models, this simplifies the right-hand side of equations 2 and 5:

$$R^2_{resid} = 1 - \frac{\hat{\sigma}_e^2}{\text{var}(Y)}$$

$$R^2_{resid.partial.x} = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_{e,r}^2} \quad (6)$$

To illustrate the difference between marginal and partial R^2 s, consider the case of the model from equation (1) in which, in addition to Y , the values of x are associated with taxon (i.e., x has taxonomic signal). Specifically, let $x = b_x + e_x$ where b_x is a random effect for taxa that has the same structure as the random variable b for Y in equation (1). To generate a contrast between marginal and partial R^2 s, let x have either weak ($\sigma_{b_x}^2 = 0.2, \sigma_{e_x}^2 = 0.8$) or strong ($\sigma_{b_x}^2 = 0.8, \sigma_{e_x}^2 = 0.2$) taxonomic signal (Table 1). The marginal values of R^2_{glmm} ($R^2_{glmm(m)}$ and $R^2_{glmm(r)}$ from equations 3 and 4) are little affected by the strength of taxonomic signal in x . In contrast, the partial R^2 values for x from $R^2_{glmm.partial}$ (equation 5) and R^2_{resid} (equation 6) are much lower for the case when there is strong taxonomic signal in x ; this is also true for R^2_{pred} and R^2_{lik} which are introduced later.

The pattern seen in the R^2 s corresponds to the statistical significance of the coefficient for x in the model: the average t -scores for the fixed effect when there was weak and strong taxonomic signal in x were 5.23 and 1.61,

respectively, which correspond to $P < 10^{-5}$ and $P = 0.11$. Thus, the lower significance of the fixed effect with strong taxonomic signal in x is reflected in the lower partial R^2 s. Note finally that $R^2_{glmm.partial}$ and R^2_{resid} have very similar values, implying that in this case the sum of all variance estimates used in equation (5) are close to $\text{var}(Y)$ in equation (6).

The explanation for taxonomic signal in x affecting partial R^2 s is that, in the reduced model with x removed, the taxonomic signal in values of Y that was accounted for by x in the full model is absorbed into the random effect b (equation 1) in the reduced model. This can be seen by looking at the taxonomic effect model [full model: $(\beta_1, \sigma_b^2) = (0, 1)$; reduced model: $(\beta_1, \sigma_b^2) = (0, 0)$]. For the case with strong taxonomic signal in x , the total R^2 for the taxonomic effect model (0.36) is similar to the total R^2 for the full model (0.38), implying that the taxonomic signal in the residuals can explain most of the taxonomic signal in the data Y .

A PGLS model comparable to the LMM in equation (1) is

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma^2 \Sigma(\theta)) \quad (7)$$

The error term e_i has a multivariate Gaussian distribution with mean 0 and covariance matrix $\sigma^2 \Sigma(\theta)$ that may depend on a vector of parameters θ . In PGLS, the structure of $\sigma^2 \Sigma(\theta)$ is typically generated under a specific model of evolution, such as an Ornstein–Uhlenbeck model of evolution under selection to an optimum (Martins and Hansen, 1997; Blomberg et al., 2003; Lavin et al., 2008). The definitions of R^2_{resid} , R^2_{pred} and R^2_{lik} are not limited by the form of evolutionary transform, although their implementation for phylogenetic models for discrete data is limited by the availability of methods for fitting discrete phylogenetic models. I will present the R^2 s using Pagel’s λ branch-length transform (Pagel, 1997; Housworth et al., 2004) that makes implementation easiest. In this transform, $\Sigma(\lambda)$ is the sum of two matrices, $\Sigma(\lambda) = \lambda \Sigma_{BM} + (1 - \lambda) \mathbf{I}$, where Σ_{BM} is the covariance matrix derived under the assumption of Brownian motion evolution (Felsenstein, 1985; Grafen, 1989), and \mathbf{I} is the identity matrix. If $\lambda = 1$, then the covariance between errors at two tips on the phylogenetic tree is proportional to the height of their most common ancestor. If $\lambda = 0$, the errors are uncorrelated, and $0 < \lambda < 1$ gives intermediate levels of phylogenetic signal. The effect of λ on the covariances among errors can be depicted by lengthening the tip branches of the BM phylogenetic tree to make up a proportion λ of the base-to-tip distance (Fig. 1, PGLS). Although the phylogenetic tree in Figure 1 is ultrametric, the calculations of R^2 s are the same for trees with non-contemporaneous tips.

The LMM in equation (1) has the same statistical structure as equation (7), since the covariance matrix for

TABLE 1. Illustrative simulation comparing R^2 s fitted to two LMMs (equation 1) that differ in whether the predictor variable x shows weak taxonomic signal ($\sigma_{bx}^2=0.2$) or strong taxonomic signal ($\sigma_{bx}^2=0.8$)

	Full	Reduced	$R^2_{\text{glmm,partial}}$	R^2_{resid}	R^2_{pred}	R^2_{lik}	$R^2_{\text{glmm(c)}}$	$R^2_{\text{glmm(m)}}$	$R^2_{\text{glmm(r)}}$
$\sigma_{bx}^2=0.2$	(β_1, σ_b^2)								
Total tax.	(0,1)	(0,0)	0.29	0.29	0.34	0.16			
Partial x	(1,1)	(0,1)	0.25	0.26	0.26	0.25			
Total	(1,1)	(0,0)	0.48	0.47	0.52	0.38	0.48	0.23	0.24
Partial tax.	(1,1)	(1,0)	0.30	0.31	0.36	0.19			
Total x	(1,0)	(0,0)	0.24	0.24	0.24	0.24			
$\sigma_{bx}^2=0.8$									
Total tax.	(0,1)	(0,0)	0.36	0.36	0.41	0.22			
Partial x	(1,1)	(0,1)	0.04	0.04	0.02	0.07			
Total	(1,1)	(0,0)	0.38	0.38	0.42	0.28	0.38	0.18	0.21
Partial tax.	(1,1)	(1,0)	0.26	0.25	0.31	0.13			
Total x	(1,0)	(0,0)	0.17	0.17	0.17	0.17			

Partial R^2 s were computed for all nested models, ordered in rows with the total R^2 in the middle. Full and reduced models are specified by the presence or absence of the two parameters (β_1, σ_b^2) . $R^2_{\text{glmm,partial}}$ is the partial $R^2_{\text{glmm(c)}}$ (equation 5). R^2_{glmm} was computed for the total R^2 ($R^2_{\text{glmm(c)}}$, equation 2) and marginal R^2 s ($R^2_{\text{glmm(m)}}$, equation 3; $R^2_{\text{glmm(r)}}$, equation 4); because $R^2_{\text{glmm(m)}}$ and $R^2_{\text{glmm(r)}}$ are computed from the full model, they are placed on the same row as the total R^2 s from the other methods. Simulations were performed with $\beta_1=0$ or 1, with eight levels of b ($\sigma_b^2=0$ or 1) and 10 observations per level ($\sigma_e^2=1.5$). The mean t -scores for the fixed effect when there was weak and strong taxonomic signal in x were 5.23 and 1.61, respectively, which correspond to $P < 10^{-5}$ and $P=0.11$. Although the values reported are from single simulations, they are close to the means obtained from 1000 simulations.

the LMM can be written $\sigma_e^2 \Sigma(\sigma_b^2) = \sigma_b^2 \Sigma_b + \sigma_e^2 \mathbf{I}$, where Σ_b is a block-diagonal matrix whose values are 1 for each row i and column j corresponding to species within the same taxon and 0 elsewhere (Gelman and Hill, 2007). Although random effects in LMMs are often explained as if they were randomly varying fixed effects, technically they occur in the variance-covariance matrix of the error term. Nonetheless, in most fitting algorithms, such as that used in lmer in the lme4 R package (Bates et al., 2015), the values of b_i at each level of the random effect are computed; technically, these values are the conditional expectations of b_i as opposed to fitted coefficients. LMMs can be fit as if they were phylogenetic models, which gives identical results as fitting them as LMMs (Ives, 2018).

For the phylogenetic model with Pagel's λ transformation, a possible approach to defining an R^2 is to treat $(1-\lambda)\sigma^2$ as the unexplained variance as done for σ_e^2 in the LMM (equation 6). This approach fails, however. The problem is that the evolutionary null hypothesis is BM phylogenetic signal ($\lambda=1$), and it is common for estimates of $\lambda=1$ in data sets. When this occurs, the total $R^2_{\text{resid}} = 1$ (equation 6) and consequently, all partial R^2 s will be 1 or undefined (0 divided by 0). This problem does not occur for the LMM because R^2_{resid} only equals 1 if there is no variation in Y among species within the same taxa; while this is not impossible, it will essentially never occur. An exception to this is models that incorporate measurement error, in which tip lengths can go to zero when estimates of measurement error in Y are too high (Ives et al., 2007); R^2_{resid} , R^2_{pred} , and R^2_{lik} can be extended for measurement-error models, although this is not described here. The problem of defining the unexplained (residual) variation needed to compute R^2_{resid} (equation 6) is more clear for other branch-length

transforms, such as the OU transform, in which $\sigma^2 \Sigma(\theta)$ does not decompose into linear terms as it does in Pagel's λ transformation.

A solution to this problem involves scaling $\Sigma\theta$. Because the total variance in the model, $\sigma^2 \Sigma(\theta)$, is the product of σ^2 and $\Sigma(\theta)$, the matrix $\Sigma(\theta)$ can be rescaled by a constant without changing the fit of the statistical model; the only effect of multiplying $\Sigma(\theta)$ by a constant is to change the value of σ^2 by $1/\text{constant}$ in the fitted model. Even though this does not affect model fitting, it does affect an R^2 that is based on σ^2 . In general $\Sigma(\hat{\theta}_f) \neq \Sigma(\hat{\theta})$, so $\Sigma(\theta)$ does not cancel out when dividing $(\hat{\sigma}_e)^2$ by $(\hat{\sigma}_{e,r})^2$ (equation 6). To solve this problem, I propose scaling $\Sigma(\theta)$ so that the sum of branch lengths of the corresponding phylogenetic tree equals 1. For a fitted tree with strong phylogenetic signal, scaling $\Sigma(\theta)$ to have the sum of branch lengths equal to 1 will make the base-to-tip distances greater than a fitted tree with no phylogenetic signal. Because this scaling increases the diagonal elements in $\Sigma(\theta)$ when there is greater phylogenetic signal, it will reduce the estimates of σ^2 and decrease the variance in the residuals that is unexplained by the model. Although this is only a convention (as opposed to a scaling derived from theory), the resulting R^2_{resid} performs for phylogenetic models in a similar way as it does for LMMs.

To illustrate R^2_{resid} for a phylogenetic model, suppose a researcher has data on sprint speed in lizards, and the predictor variables are hind leg length and body size (e.g., Bauwens et al., 1995). Hind leg length is the variable of interest, while body size is a nuisance variable, and therefore, the interesting R^2 compares the models with and without hind leg length. I simulated the case of 30 species under BM evolution and compared the cases in which hind leg length (as a proportion of

TABLE 2. Simulation example of sprint speed regressed on log body mass (x_1) and log hind limb length (x_2)

Coefficient	Phylogenetic signal in x_2		No phylogenetic signal in x_2	
	Estimate	P	Estimate	P
Intercept	0.34	0.58	0.28	0.65
x_1	1.17	$<10^{-5}$	1.00	$<10^{-5}$
x_2	0.21	0.22	0.20	0.0038
λ	1		1	
$N = 30$	logLik = -16.92		logLik = -16.96	

Reduced model	Phylogenetic signal in x_2			No phylogenetic signal in x_2		
	R^2_{resid}	R^2_{pred}	R^2_{lik}	R^2_{resid}	R^2_{pred}	R^2_{lik}
$x_2 = 0$	0.05	-0.05	0.05	0.16	0.22	0.26
$\lambda = 0$	0.40	0.65	0.51	0.46	0.65	0.56
$x_1 = x_2 = \lambda = 0$	0.86	0.92	0.89	0.78	0.86	0.82

For the simulation, the regression coefficients for $\log(\text{body size})$ and $\log(\text{hind leg length})$ were $\beta_1 = 1$ and $\beta_2 = 0.5$, and the intercept was $\beta_0 = 0$ (equation 8); residual variation was given by BM evolution, so $\lambda = 1$. Hind limb length was simulated under BM evolution (left table) or as a Gaussian random variable (right table), and in both cases the variance of the hind limb length was standardized to 1.

body size) either did or did not show phylogenetic signal. Specifically, the model was

$$Y = \beta_0 + \beta_1 \log(\text{body size}) + \beta_2 \log(\text{hind leg length}) + e_i$$

$$e_i \sim \text{Gaussian}(\mathbf{0}, \sigma^2 \Sigma_{BM}) \quad (8)$$

where $\log(\text{body size})$ was selected from a Gaussian distribution with covariance matrix Σ_{BM} , and $\log(\text{hind leg length})$ was selected from a Gaussian distribution with covariance matrix either \mathbf{I} or Σ_{BM} . In the fitted PGLS models without and with phylogenetic signal in hind leg length (Table 2), the parameter estimates and log likelihoods were similar, and the only indication that phylogenetic signal in hind limb length affected the fit of the model was the P -value for the regression coefficient for hind limb length ($P = 0.22$ and $P = 0.0038$ for the models with and without phylogenetic signal in hind leg length). However, the partial R^2_{resid} (and also partial R^2_{pred} and R^2_{lik} included for completeness) for hind leg length were ≤ 0.05 when hind leg length had phylogenetic signal, and 0.16–0.26 when it did not. (Note that the partial R^2_{pred} was negative for the model with phylogenetic signal in x_2 , which can happen especially with small sample sizes.) In contrast, the partial R^2_{resid} for phylogenetic signal (reduced model with $\lambda = 0$) and the total R^2_{resid} (reduced model with $x_1 = x_2 = \lambda = 0$) did not differ as much between simulations. The explanation for these results is similar to that for the example with the LMM (Table 1). The partial R^2 s for hind limb length depended upon the phylogenetic signal in hind limb length because, when there is phylogenetic signal and hind limb length is removed from the model, much of the information is recaptured in the phylogenetic signal of the residual variation. This example illustrates the value of having partial R^2 s that can assess the role of predictor variables separately from other variables like body size that are not of specific interest. It also shows

how the partial R^2 for the phylogeny is associated with the statistical significance of tests for parameters like λ that depend on phylogeny.

R^2_{resid} can be extended to models of discrete data, and for illustration I will use the simple model with only a single predictor variable (fixed effect) that is comparable to equations (1) and (7):

$$E\{Y_i | e_i\} = \mu_i$$

$$g(\mu_i) = \beta_0 + \beta_1 x_i + e_i$$

$$e_i \sim \text{Gaussian}(\mathbf{0}, \sigma^2 \Sigma(\theta)) \quad (9)$$

where data Y_i are distributed by a member F of the exponential family of distributions (McCullagh and Nelder, 1989; Gelman and Hill, 2007). The parameter μ_i of distribution F is itself a random variable, and applying the link function $g()$ to μ_i gives a linear equation in terms of the predictor variable x_i and a random variable e_i . The variable e_i has a multivariate Gaussian distribution with means 0 and covariance matrix $\sigma^2 \Sigma(\theta)$ that may depend on a vector of parameters θ ; hence, the second line of equation (9) is often called the Gaussian- or latent-space equation. When the link function $g()$ is the identity function, then equation (1) becomes a linear model (e.g., LMM or PGLS). For the GLMM companion to the LMM in equation (1) with a random effect b , the covariance matrix is $\sigma_e^2 \Sigma(\sigma_b^2) = \sigma_b^2 \Sigma_b$ in which there is no variance term σ_e^2 .

For comparison with LMMs, GLMMs can also be depicted as a tree (Fig. 1 GLMM), but rather than the residual errors at the tips of the tree having length 1, for models of discrete data the error variance is generated by the unavoidable differences between the observation (0 or 1 for binary data) and probability of the observation (given by μ_i that takes values between 0 and 1). The lengths of these variances $\sigma_{w[i]}^2$ will differ among observations i , since they depend on the difference between an observation and μ_i (Schall, 1991). The terms

$\sigma_{w[i]}^2$ are the inverses of the GLM iterated weights that are used for fitting GLMs (McCullagh and Nelder, 1989) and occur in methods for fitting GLMMs (Schall, 1991; Breslow and Clayton, 1993; Pinheiro and Chao, 2006) and PGLMMs (Ives and Helmus, 2011). Specifically,

$$\sigma_{w[i]}^2 = g'(\mu_i) \text{var}(y_i - \mu_i) \tag{10}$$

where $g'(\mu_i)$ is the derivative of the link function at μ_i , and $\text{var}(y_i - \mu_i)$ denotes the theoretical variance of the difference between datum y_i and its predicted value μ_i . For example, for a logit link function in a binary GLMM, $\text{var}(y_i - \mu_i) = \mu_i(1 - \mu_i)$.

For GLMMs, R_{resid}^2 is defined exactly like $R_{glmm(c)}^2$ for GLMMs:

$$R_{resid}^2 = 1 - \frac{\sigma_d^2}{\hat{\sigma}_x^2 + \hat{\sigma}_b^2 + \sigma_d^2} \tag{11}$$

where σ_d^2 gives the “distribution-specific variance” that depends on the form of the link function $g()$. For a binomial GLMM with probit and logit link function, Nakagawa and Schielzeth (2013) propose $\sigma_d^2 = 1$ and $\sigma_d^2 = \pi^2/3$, respectively. Nakagawa et al. (2017) give an extensive discussion of additional formulations of σ_d^2 for different exponential distributions and link functions. As detailed in the Appendix, $\sigma_d^2 = \pi^2/3$ tends to underestimate R_{resid}^2 (and $R_{glmm(c)}^2$) for the logit link function relative to the probit link function. Reducing σ_d^2 to $0.8768809 \pi^2/3$ leads to a better correspondence between R_{resid}^2 calculated for the same data using probit and logit link functions.

An alternative approach for determining σ_d^2 is to use the GLM iterated weights $\sigma_{w[i]}^2$, since these are associated with the unexplained variance in the Gaussian space of the GLMM (equation 9). Because $\sigma_{w[i]}^2$ will differ among data points, it is necessary to take an average; the appropriate average is the geometric mean (Appendix). This leads to using $\sigma_d^2 = \sigma_w^2 = \exp(E\{\log \sigma_{w[i]}^2\})$. For a binary GLMM, using either $\sigma_d^2 = 0.8768809 \pi^2/3$ or $\sigma_d^2 = \sigma_w^2$ leads to the very similar values of R_{resid}^2 for both logit and probit link functions. Nonetheless, because the values of σ_d^2 differ, so do the values of R_{resid}^2 which are lower when scaling $\sigma_d^2 = \sigma_w^2$. Throughout the remainder of this article, the binary GLMM will use a logit link function with $\sigma_d^2 = 0.8768809 \pi^2/3$, which is most comparable to $R_{glmm(c)}^2$, although $\sigma_d^2 = \sigma_w^2$ is the default in the R package rr2.

R_{resid}^2 for GLMMs extends to partial R²s using the standard formula

$$R_{resid.partial}^2 = 1 - \frac{1 - R_{resid.f}^2}{1 - R_{resid.r}^2} \tag{12}$$

where $R_{resid.f}^2$ and $R_{resid.r}^2$ are R_{resid}^2 (equation 11) for full and reduced models, respectively. For phylogenetic models, R_{resid}^2 takes exactly the same form (Fig. 1, PGLMM). GLMMs can be fit using `glmer()` in the R package lme4 (Bates et al., 2015), and binary phylogenetic GLMMs (PGLMMs) (Ives and Helmus, 2011; Ives and Garland, 2014) can be fit using the function `binaryPGLMM()` available in the R packages `ape` (Paradis et al., 2004) and `rr2` (Ives and Li, 2018). In principle, it is possible to calculate R_{resid}^2 for the Bayesian methods in the R package MCMCglmm (Hadfield, 2015), although this is not pursued here.

$$R_{pred}^2$$

R_{pred}^2 (for prediction) is based on the variance in the difference between observed and predicted data, $\frac{1}{n} \sum (Y_i - \hat{Y}_i)$. This approach conceptually comes the closest to answering the question of how much variation in the data is explained by the covariances in the model, where the “explanation” is measured by the prediction of the residuals. For the case of LMMs, the predicted values \hat{Y}_i are taken as the sum of the fixed effect values and the conditional estimates of the values of the random effects. For the LMM in Figure 1, this corresponds to the estimated values at the polytomies formed at the nodes shared by all observations within the same level of the random effect (i.e., taxon). As the number of observations within each level of the random effect increases, R_{pred}^2 for LMMs converges to R_{resid}^2 because the estimates of the values of the random effects become more precise.

As it does for R_{resid}^2 , PGLS poses a complication for R_{pred}^2 : what is the predicted value \hat{Y}_i ? To parallel R_{pred}^2 for LMMs, \hat{Y}_i could be taken as the estimated value at the node immediately below the tip on the phylogeny containing Y_i . For phylogenies with some short terminal branch lengths, however, the estimates for the node underneath Y_i will be determined largely by the value of Y_i itself, leading to very high R²s. Therefore, for PGLS I define R_{pred}^2 using the estimates \hat{Y}_i computed by removing the point Y_i from the data set and then estimating \hat{Y}_i from the predictor variables and the remaining data points. The estimate of \hat{Y}_i is calculated using the full $n \times n$ covariance matrix of the fitted model, $\mathbf{V}(\hat{\theta}) = \sigma^2 \Sigma(\hat{\theta})$, that contains the variances and covariances of the residual errors. Specifically for equation (1), the expected value of residual $R_i = Y_i - (\beta_0 + \beta_1 x_i)$ from the remaining residuals $\mathbf{R}_{[-i]}$ is

$$\hat{R}_i = \bar{R} + \mathbf{V}_{[i,-i]} \mathbf{V}_{[-i,-i]}^{-1} (\mathbf{R}_{[-i]} - \bar{R}), \tag{13}$$

where \bar{R} is the GLS mean of the residuals, $\mathbf{V}_{[i,-i]}$ is row i of $\mathbf{V}(\hat{\theta})$ with column i removed, and $\mathbf{V}_{[-i,-i]}$ is $\mathbf{V}(\hat{\theta})$ with

row i and column i removed (Petersen and Pedersen, 2012). The predicted value of Y_i is then $\hat{Y}_i = \beta_0 + \beta_1 x_i + \hat{R}_i$.

The same approach as used for LMMs can be used for GLMMs and PGLMMs. In these cases, the variances are calculated for untransformed values of Y_i , rather than in the Gaussian space of $g(\mu_i)$ as was done for R^2_{resid} . Therefore, one should expect that they may take different values. The predicted values are given from the estimation algorithms for GLMMs by `glmer()` in the R package `lme4` (Bates et al., 2015) and for PGLMMs by `binaryPGLMM()` (Ives and Helmus, 2011; Ives and Garland, 2014) in the R packages `ape` (Paradis, 2012) and `rr2` (Ives and Li, 2018). In PGLMMs, the estimates \hat{Y}_i correspond to those values a distance σ_w^2 from the tips of the tree in Figure 1.

$$R^2_{lik}$$

Motivation for R^2_{lik} (for likelihood) comes from the standard R^2 for a generalized least squares model, GLS (Buse, 1973; Judge et al., 1985),

$$R^2 = 1 - \frac{SS_f(\hat{\theta}_f)}{SS_r(\hat{\theta}_r)} \quad (14)$$

where SS_f and SS_r are the GLS sums-of-squared errors for the full and reduced models (Judge et al., 1985):

$$SS(\hat{\theta}) = (\mathbf{Y} - \mathbf{X}\hat{\beta})' \mathbf{V}(\hat{\theta})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta}). \quad (15)$$

Here, \mathbf{Y} is the $n \times 1$ vector of response values Y_i , \mathbf{X} is the $n \times p$ matrix for p predictor variables (including the intercept), and $\hat{\beta}$ is the $1 \times p$ vector of estimated regression coefficients (fixed effects). Technically, GLS models do not have covariances that depend on parameters θ (even though "PGLS" is used to refer to models that do), so the SS in equation (15) is conditioned on θ .

For continuous phylogenetic models, equation (14) presents the same problem as discussed for R^2_{resid} : comparing full and reduced models depends on the scaling of $\Sigma(\theta)$, because $\mathbf{V}(\theta) = \sigma^2 \Sigma(\theta)$. For R^2_{resid} this problem was solved by the convention of standardizing $\Sigma(\theta)$ to have summed branch lengths equal to one. Here, a different scaling is used. In particular, if $\Sigma(\theta)$ is scaled so that the determinant of $\mathbf{V}(\theta)$ equals one, then

$$\log Lik(\hat{\theta}) = -\frac{n}{2} (\log(2\pi SS(\hat{\theta})/n) + 1). \quad (16)$$

Substituting into equation (14) then leads to

$$R^2_{lik} = 1 - \exp\left(\frac{-2}{n} (\log Lik(\hat{\theta}_f) - \log Lik(\hat{\theta}))\right). \quad (17)$$

I derived R^2_{lik} defined by equation (17) for PGLS models, although it can be extended to any model that is fit using maximum likelihood. R^2_{lik} has been proposed for logistic regression (Cragg and Uhler, 1970; Maddala, 1983; Cox and Snell, 1989) and generalized by Magee

(1990) and Pinheiro and Chao (2006), and used for LMMs by Kramer (2005). For discrete data, equation (17) does not have a maximum of one, because the maximum attainable log-likelihood for discrete data is zero. Therefore, Nagelkerke (1991) and Cameron and Windmeijer (1996) proposed dividing by the maximum attainable value, which is equation (17) with $\log Lik(\hat{\theta}_f) = 0$; throughout, I have used this standardization.

R^2_{lik} is computing using maximum likelihood (rather than REML) estimates for LMMs, PGLS, and GLMM models. A complication occurs for binary PGLMM models, because the algorithm used by `binaryPGLMM()` from which R^2_{resid} and R^2_{pred} are calculated uses quasi-likelihoods and does not give a true maximum likelihood. Therefore, for binary phylogenetic models I calculated R^2_{lik} using the ML-based function `phyloglm()` in the R package `phylolm` (Ho and Ane, 2014), fitting the model with penalized ML but then using the provided log likelihood values to calculate R^2_{lik} . The function `phyloglm()` fits a phylogenetic logistic regression (Ives and Garland, 2010), rather than the PGLMM given by equation (9), although I will still refer to this as a binary PGLMM model for simplicity in the text.

Simulations for Assessment

The simulations to assess the statistical properties of the R^2 s applied to LMM, PGLS, GLMM, and PGLMM all follow the same strategy. For each, data were simulated for the cases with variation in a predictor variable and no covariances ($\beta_1 > 0, \theta = 0$), only covariances ($\beta_1 = 0, \theta > 0$) or both ($\beta_1 > 0, \theta > 0$). For each case, the model parameters were the same for all simulations, so that variation in values of a given R^2 among data sets is caused by random sampling from the same statistical process.

For LMM, data were simulated with the model

$$y_i = \beta_0 + \beta_1 x_i + b u_i + e_i \quad (18)$$

where x_i follows a Gaussian distribution with mean 0 and variance 1, and the random effect u_i has 10 levels, with b following a Gaussian distribution with mean 0 and variance θ . I selected parameter values to generate moderate R^2 values. For GLMMs, values from equation (18) without the residual error term e_i were used through a logit link function (equation 9) to produce binomial probabilities for a binary model.

For the PGLS model, to obtain the covariance matrix $\Sigma(\theta)$ in equation (7), I first simulated random phylogenetic trees using the `rtree()` function of the `ape` package of R (Paradis et al., 2004). Thus, a different tree was simulated for each data set. The strength of phylogenetic signal was varied using Pagel's λ transformation. Values of x_i were simulated under the BM assumption using the `rTraitCont()` function (Paradis et al., 2004). The simulated data were fit using penalized maximum likelihood

with the function `phylolm()` assuming a Pagel's lambda transformation in the package `phylolm` in R (Ho and Ane, 2014). The PGLMM was similar to the PGLS model, but in contrast the predictor variable x_i was assumed to be independently distributed; including phylogenetic signal in x_i caused challenges for model fitting for some simulated data sets, making the simulation studies difficult. Phylogenetic signal in the residuals e_i was controlled by setting $\Sigma(\theta) = \theta \Sigma_{BM}$ so that in the absence of phylogenetic signal ($\theta = 0$) the simulations conformed to simple logistic regression. To simulate binary data, a logit link function was used.

RESULTS

The R^2 s were assessed according to three properties: (i) their ability to measure goodness-of-fit as benchmarked by the LLR of full model and the model with only an intercept, (ii) whether they can identify separate sources of variation in the model, and (iii) the precision of their inference about goodness-of-fit. Property (iii) treats the R^2 s as if they were statistical estimators of goodness-of-fit and asks how variable are the estimates when applied to repeated simulations from the same model (e.g., Cameron and Windmeijer, 1996). A more comprehensive assessment is given in [Supplementary Section 1](#) available on Dryad at <http://dx.doi.org/10.5061/dryad.345v6>.

Goodness-of-Fit

All R^2 s were positively related to the LLR (Fig. 2), which is a minimum requirement for an R^2 . R^2_{lik} shows a monotonic relationship with LLR, which is necessarily the case due to the definition of R^2_{lik} (equation 17). For the remaining R^2 s, values for a given LLR were generally lower for simulations in which variation was produced only by the fixed effect ($\beta_1 > 0, \theta = 0$; Fig. 2, circles). This implies that, relative to the LLR, these R^2 s were attributing less "explained" variance to fixed effects than random effects.

For the LMM, I included the adjusted R^2 , R^2_{adj} , computed from OLS regression by treating the random effect as a categorical fixed effect. R^2_{resid} and R^2_{adj} were almost identical. This correspondence implies that R^2_{resid} gives an R^2 that is interpretable in the same way as the standard R^2_{adj} but generalized to LMMs.

All of the R^2 s other than R^2_{lik} showed greater scatter in their relationships with LLR for the simulations of binary data (GLMM and PGLMM). In part, this is due to the difficulty of estimating variance parameters θ in binomial models. The scatter seems particularly large for R^2_{resid} and R^2_{pred} applied to PGLMM simulations, although this case requires some technical discussion. For PGLMM, the LLR was obtained from `phylglm()`

using penalized maximum likelihood, whereas R^2_{resid} and R^2_{pred} were estimated from the model fit by `binaryPGLMM()` using the pseudo-likelihood. The `phylglm()` estimate of phylogenetic signal, λ , tended to absorb at zero even when the estimate λ from `binaryPGLMM()` was positive; therefore, R^2_{resid} and R^2_{pred} could be positive even when the LLR was zero. Previous comparison between `phylglm()` and `binaryPGLMM()` showed that they have similar performances but do not necessarily give the same conclusions about the presence of phylogenetic signal for the same data set (Ives and Garland, 2014).

Identifying Sources of Variation

The partial R^2_{resid} , R^2_{pred} , and R^2_{lik} were generally able to identify sources of variation between components of a model, specifically between regression coefficients (fixed effects) and covariance parameters (random effects); this is seen in the generally non-overlapping clustering of symbols in Figure 3. Simulations with $\beta_1 > 0$ and $\theta = 0$ should have partial R^2 s for β_1 that are positive and partial R^2 s for θ that are zero (circles, Fig. 3). Simulations with $\beta_1 = 0$ and $\theta > 0$ should have partial R^2 s for β_1 that are zero and partial R^2 s for θ that are positive (triangles, Fig. 3). Simulations with $\beta_1 > 0$ and $\theta > 0$ should have both partial R^2 s positive (x 's, Fig. 3). Because in the simulations the values of β_1 and θ were the same whether or not the other was zero, the partial R^2 s for β_1 should be the same for simulations with $\theta = 0$ (circles) as for simulations with $\theta > 0$ (x 's), and the partial R^2 s for θ should similarly be the same for $\beta_1 = 0$ (triangles) and $\beta_1 > 0$. For continuous data (LMM and PGLS), all three R^2 s had similar performance and similar values of the partial R^2 s (see also [Supplementary Section 1](#) available on Dryad). For binary data (GLMM and PGLMM), the three R^2 s showed more scatter, which in large part is due to the greater statistical challenge of estimating regression coefficients and variance parameters from discrete data. This is seen, for example, in the GLMM and PGLMM simulations with $\beta_1 > 0$ and $\theta > 0$ in which the partial R^2_{lik} for θ was occasionally zero (x 's); these cases occur when the estimate of θ was zero even though a non-zero value was used in the simulations.

Inference about the Underlying Process

The ability of R^2 s to infer the fit of a model to its assumed statistical process depends on the precision of the estimates of R^2 . Figure 4 plots the mean values of the R^2 s with 66% and 95% inclusion intervals for simulated data sets with sample sizes 40, 60, ..., 160. For LMMs and GLMMs, there were 10 levels of the random effect; data sets were produced by first simulating 160 samples (16

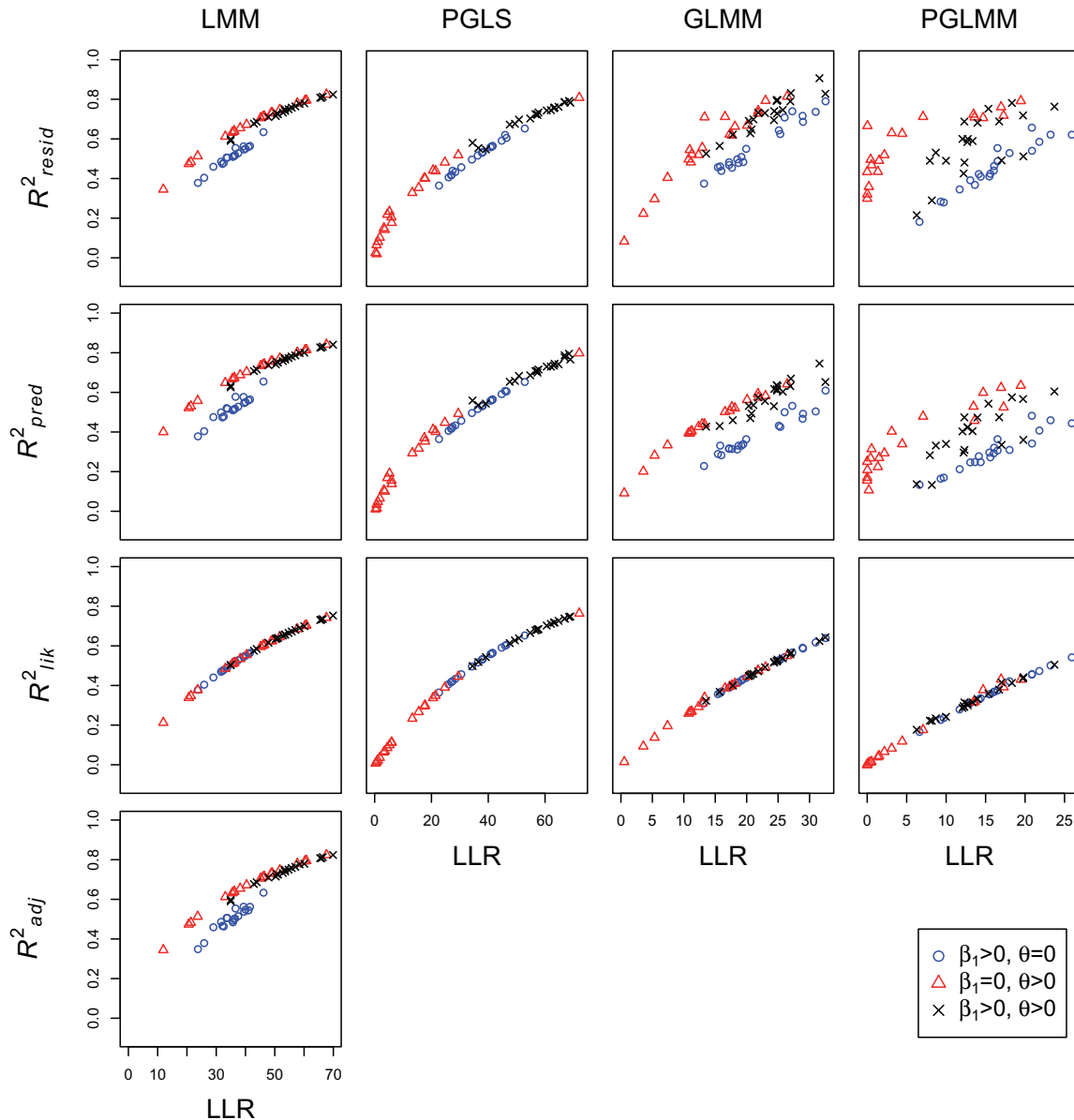


FIGURE 2. Results for LMM, PGLS, GLMM, and PGLMM simulations giving R^2_{resid} , R^2_{pred} , R^2_{lik} , and the OLS R^2_{adj} versus the LLR between full model and reduced model containing only an intercept. All simulated data had 100 samples. For LMM, the simulation model (equation 18) contained a fixed effect with $\beta_1 = 0$ or 1, and a random effect u_i with 10 levels and variance $\theta = 0$ or 1.5. The binomial (binary) GLMM was similar but with $\beta_1 = 0$ or 1.8, and $\theta = 0$ or 1.8. For PGLS, $\beta_1 = 0$ or 1, and the strength of phylogenetic signal $\theta = \lambda = 0$ or 0.5; while for PGLMM $\beta_1 = 0$ or 1.5, and $\theta = 0$ or 2. The LMM was fit using `lmer()` (Bates et al., 2015); the GLMM was fit using `glmer()` (Bates et al., 2015); the PGLS was fit using `phylolm()` (Ho and Ane, 2014); and for PGLMM LLR and R^2_{lik} were fit using `phyloglm()` (Ho and Ane, 2014), and R^2_{resid} and R^2_{pred} were fit using `binaryPGLMM()` (Ives and Garland, 2014). For reduced models without variance parameters, fitting was done with `lm()` and `glm()` in R.

replicates at each level) and then randomly removing two replicates at each level to reduce the sample size in steps of 20. For PGLS and PGLMM, each data set at each sample size was simulated independently.

For LMM simulations, R^2_{resid} , R^2_{pred} , and R^2_{adj} showed similar patterns (Fig. 4), reflecting the fact that they give very similar values (Fig. 2, Supplementary Section 2 available on Dryad). Mean values did not change

much with sample size, and there was only a moderate increase in variability among simulations with decreasing sample size. In contrast, mean values of R^2_{lik} decreased with decreasing sample size. This probably reflects that information is lost when estimating the model parameters. In contrast to LMM simulations, the PGLS simulations showed less change in the means of R^2_{lik} with sample size, presumably because there were

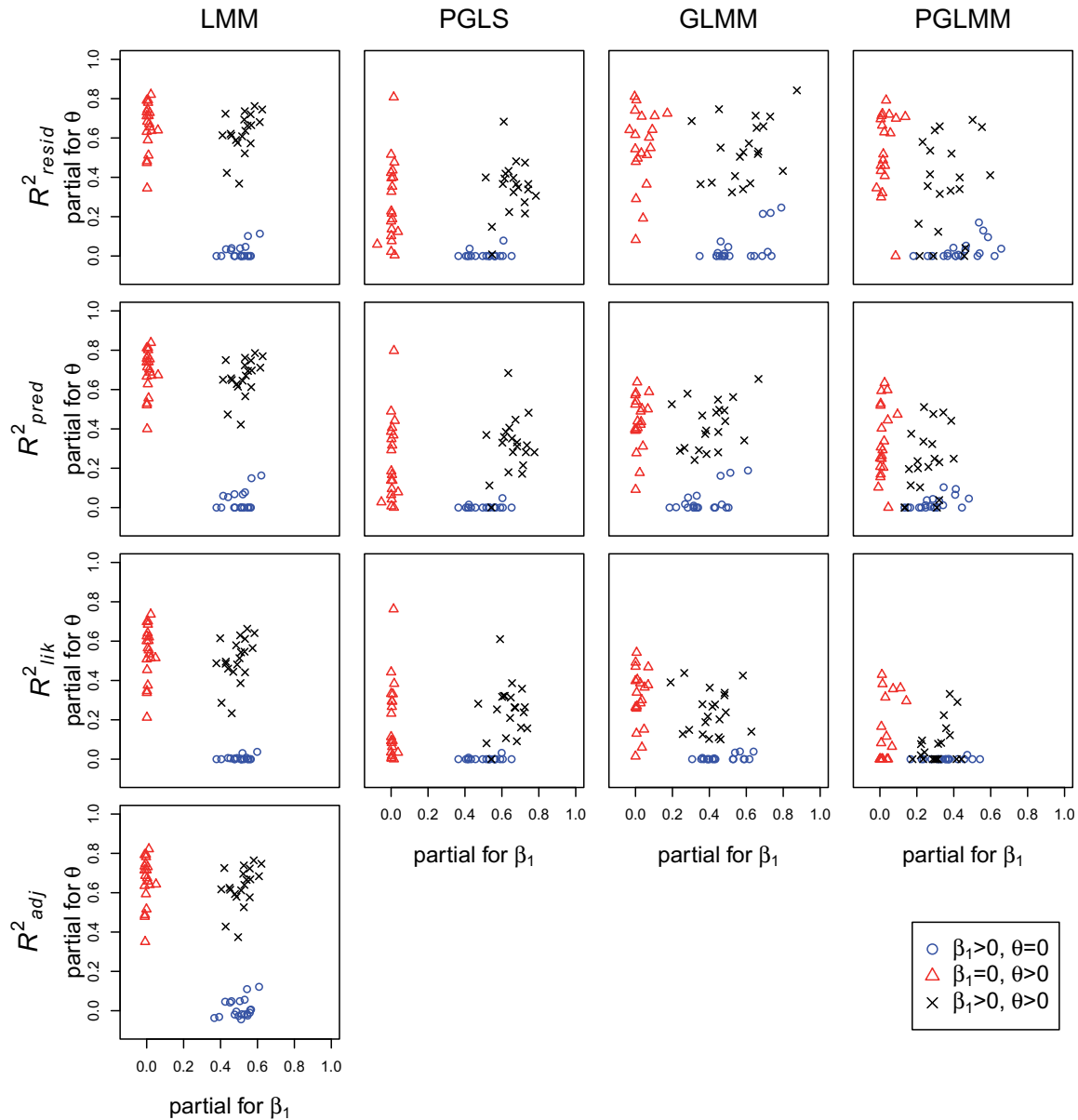


FIGURE 3. Results for LMM, PGLS, GLMM, and PGLMM simulations giving partial values of R^2_{resid} , R^2_{pred} , R^2_{lik} , and R^2_{adj} . The partial R^2 for β_1 was calculated using the reduced model in which θ is removed, and for the partial R^2 for θ the reduced model had β_1 removed. The simulated data and fitting methods are the same as in Figure 2.

more covariances among samples (i.e., the covariance matrix had more non-zero elements) than in the LMM with few replicates per level.

For the GLMM, both R^2_{resid} and R^2_{pred} had somewhat higher variances (less precision) than R^2_{lik} . The greater variation in values of R^2_{resid} and R^2_{pred} compared with R^2_{lik} may occur because R^2_{resid} and R^2_{pred} depend directly on estimates from the models, whereas R^2_{lik} depends only on the synoptic likelihood. Thus, R^2_{resid} and R^2_{pred} are compromised when the estimates are poor, as is

particularly the case when sample sizes are small. For PGLMM, the variances in R^2_{resid} and R^2_{pred} were similar to R^2_{lik} (Fig. 4). This is likely because estimates of phylogenetic signal ($\lambda = \theta$) were well-bounded, in contrast to the variance in random effects in the GLMMs.

DISCUSSION

R^2_{resid} , R^2_{pred} , and R^2_{lik} were presented here with focus on phylogenetic models, although they are broadly applicable to models with correlated errors. Below, I first address their use as measures of goodness-of-fit, and

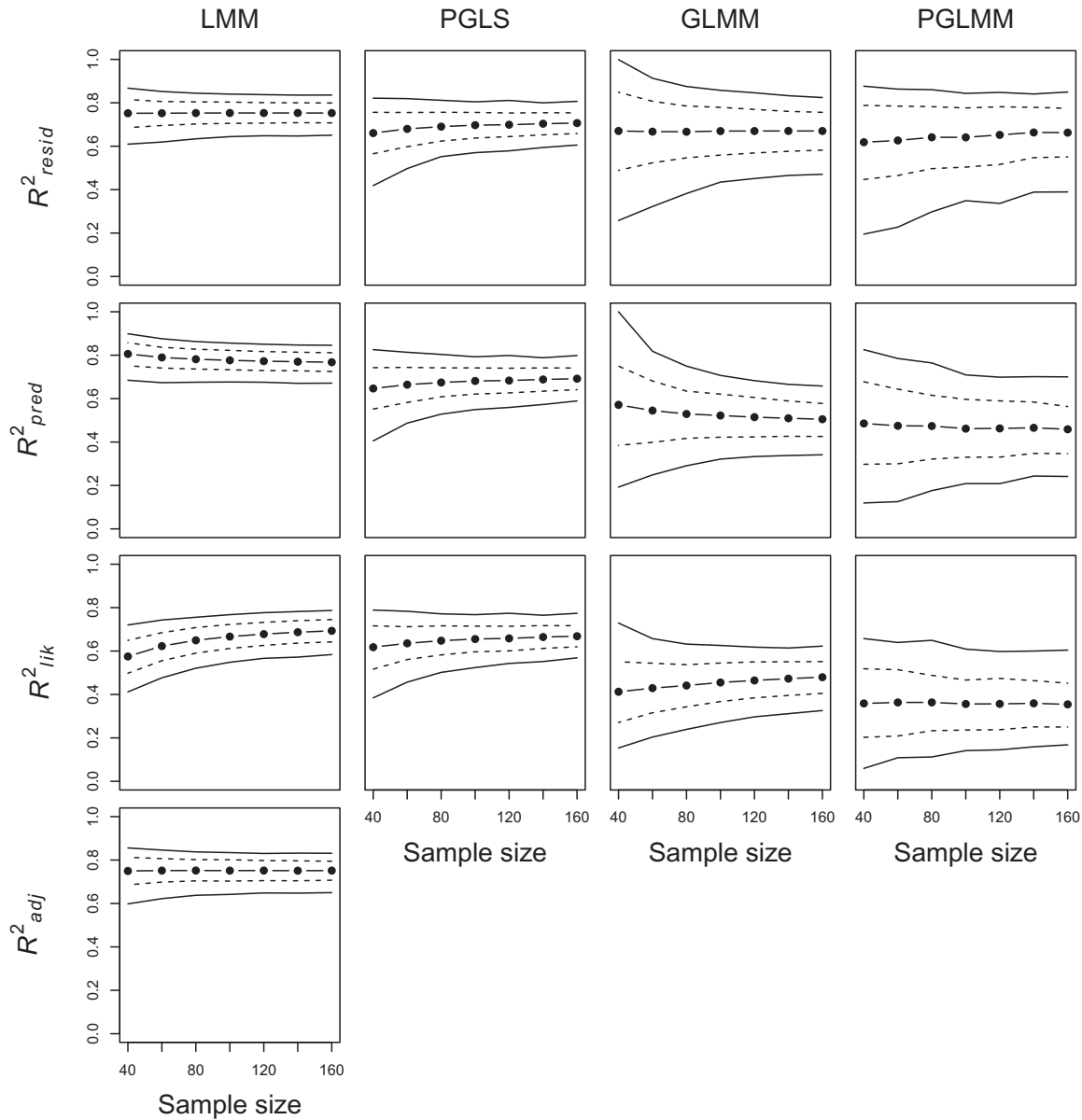


FIGURE 4. Results for LMM, PGLS, GLMM, and PGLMM simulations showing means, 66% and 95% inclusion intervals for R^2_{resid} , R^2_{pred} , R^2_{lik} , and R^2_{adj} versus sample size. For all simulations 1000 data sets were analyzed at each sample size. Parameter values were: LMM, $\beta_1 = 1, \theta = 1.5$; PGLS, $\beta_1 = 1, \theta = 0.5$; GLMM, $\beta_1 = 1.8, \theta = 1.8$; and PGLMM, $\beta_1 = 1.5, \theta = 2$.

then their specific application to phylogenetic models using mixed models as a reference. I end with general recommendations.

Partial R^2 and Goodness-of-Fit

The R^2 s presented here are designed to give partial R^2 s and are consequently measures of goodness-of-fit. Applications of standard R^2 s in regression and ANOVA all rely on partial R^2 s which compare the amount of variation explained by a full model to a reduced model with terms from the full model removed. Comparisons

between full and reduced models underlie model selection, either by classical methods such as stepwise selection based on partial F scores or methods involving model selection criteria such as AIC. Partial R^2 s are also closely associated with the statistical significance of a component of a model; this is especially the case for R^2_{lik} that is computed from log-likelihoods, and therefore, is directly related to likelihood ratio tests.

The partial R^2_{resid} , R^2_{pred} , and R^2_{lik} differ conceptually from the marginal $R^2_{glmm(r)}$ (equation 3) derived for GLMMs (Nakagawa and Schielzeth, 2013). While marginal R^2 s are useful in partitioning variances among

components of fitted models, partial R^2 s answer the question of how important are the components, where importance is measured by the loss of explanatory power following their removal. Because marginal R^2 s do not compare between full and reduced models, they do not give information about how the fit of a model declines as components are removed. A final advantage of partial R^2 s is that they make it easy to investigate any combination of fixed and random effects, and their interactions, by comparing full and reduced models; in contrast, computing and interpreting marginal R^2 s for complex random effects, such as random slope effects, is less straightforward (Johnson, 2014).

Partial R^2 s can answer the question of how important phylogenies are in explaining the pattern of covariance in the data. Does this mean that partial R^2 s themselves are measures of phylogenetic signal (sensu Blomberg et al., 2003) or, equivalently, phylogenetic heritability (Lynch, 1991)? Phylogenetic signal could be computed, for example, by comparing the reduced model containing no phylogenetic term (e.g., with Pagel's $\lambda = 0$) with the full model including phylogeny. However, in the phylogenetic literature, the term "phylogenetic signal" is associated with the magnitude of phylogenetic correlations in the branch-length transform (e.g., Pagel's λ) that is estimated during model fitting (Blomberg et al., 2003; Revell et al., 2008). Phylogenetic signal is akin to the marginal $R^2_{glmm(r)}$ which is based on partitioning variances into a correlated component (due to random effects) and uncorrelated residuals. Phylogenetic signal is thus a description of the phylogenetic correlations, rather than a measure of how important are the phylogenetic correlations in the goodness-of-fit of the model. Therefore, I think it is best to make a clear conceptual difference between partial R^2 s giving information about the statistical significance of phylogenetic correlations in the model, and phylogenetic signal that gives a measure of their magnitude.

Applications to LMM, PGLS, GLMM, and PGLMM

For both continuous data (LMM and PGLS) and discrete data (GLMM and PGLMM), all R^2 s had good performance, without one standing out as obviously the best. For the simple model with a single regression coefficient β_1 and single covariance parameter θ (equation 18), all R^2 s were reasonable measures of goodness-of-fit, as assessed against the LLR between full and reduced models (Fig. 2). Nonetheless, R^2_{resid} and R^2_{pred} gave lower R^2 values for variation explained by fixed effects (β_1) than random effects (θ) in comparison to R^2_{lik} and the LLR (Fig. 2). Also, although all three R^2 s gave very similar values for the same data set with continuous data, the values differed more for discrete data fit with either GLMM or PGLMM (Fig. 2 and Supplementary Section 1 available on Dryad). This is reflected in general by

the decreased precision of the R^2 s applied to discrete data, as measured by the variation in values when fit to data simulated under the same parameter values (Fig. 4). All R^2 s were capable of identifying whether β_1 or θ was responsible for the fit of the model to the data as determined by the partial R^2 s; when β_1 or θ were zero in the simulations, the partial R^2 s for β_1 or θ , respectively, were low (Fig. 3). However, the partial R^2 s for GLMM and PGLMM tended to be more variable and less conclusive than for LMM and PGLS (Fig. 3). Also, partial R^2_{resid} and R^2_{pred} can sometimes be negative, but negative values of R^2_{lik} are not possible as long as the full and reduced models are fitted correctly. Finally, R^2_{lik} decreased as sample sizes decreased especially for LMMs but also for GLMMs (Fig. 4). This is an understandable consequence of the loss of information to separate full and reduced models when there are fewer data. Nonetheless, it is an undesirable property when comparing across data sets with different sample sizes.

The poorer performance of all three R^2 s for GLMM and PGLMM relative to LMM and PGLS in terms of identifying sources of variation (Fig. 3) and precision (Fig. 4) is due to the greater challenges of fitting discrete data. This will affect the R^2 s differently if they are differently sensitive to the fitting. R^2_{resid} is calculated from fitted variances in a model; R^2_{pred} is calculated from the fitted values of Y_i ; and R^2_{lik} is calculated from the likelihood. Therefore, the three R^2 s will be sensitive to the precision with which each of these attributes is estimated. For GLMMs, the precision of R^2_{lik} was slightly greater than the other two R^2 s, although this did not appear to be the case for PGLMM. Nonetheless, for GLMMs and PGLMMs R^2_{resid} and R^2_{pred} can have negative values especially for small sample sizes, while this is never the case for R^2_{lik} .

Recommendations

An ideal R^2 would make it possible to compare among different models and among different methods used to fit the same model (Kvalseth, 1985 properties of a good R^2 #4 and #5). R^2_{resid} and R^2_{pred} can be used for any model and fitting method that estimates the covariance matrix (R^2_{resid}) and/or fitted values (R^2_{pred}); for example, they could be used to compare LMMs fit with ML vs. REML, or binary phylogenetic models fit with ML (e.g., phyloglm; Ho and Ane, 2014) or quasi-likelihood (e.g., binaryPGLMM; (Ives and Garland, 2014)). Nonetheless, R^2_{resid} and R^2_{pred} have a disadvantage in terms of generality. For R^2_{resid} a decision must be made about how to weight the covariance matrix $\Sigma(\theta)$ (equation 9), and for R^2_{pred} a decision has to be made

about how values of Y_i are predicted. In contrast, R_{lik}^2 is restricted to models that are fit with ML estimation; however, if ML is used for fitting, then values of R_{lik}^2 can be compared across different types of models. This applies to any type of data and model fit with ML estimation.

An ideal R^2 should also be intuitive (Kvalseth, 1985 property #1). However, intuitive is in the eye of the beholder. R_{resid}^2 is similar to the OLS R^2 , which grounds R_{resid}^2 in the familiar and intuitive OLS framework. R_{lik}^2 is also related to the OLS R^2 : for LMMs and PGLS, R_{lik}^2 is the same as the GLS R^2 if the determinant of $\sigma^2 \Sigma(\theta)$ is scaled to equal one. R_{pred}^2 predicts the data from covariances estimated in the model, and therefore, it is the most intuitive way to relate the variance explained by regression coefficients (fixed effects) to that explained by variance parameters (random effects). I think this is the most intuitive R^2 . Therefore, if the question is how much variation in the data is explained by a model, or different components of a model, then I think R_{pred}^2 gives the most direct answer.

An ideal R^2 would make it possible to compare the fit of one model to one data set, and the same model to another data set; in other words, it should give an overall sense of how well a model fits. For this, R_{resid}^2 and R_{pred}^2 have the advantage over R_{lik}^2 in being relatively insensitive to sample size (Fig. 4). Because R_{lik}^2 is based on the likelihood ratio between full and reduced models, all else being equal, it will decrease for smaller data sets because there is less statistical power. If some data sets are very small, R_{resid}^2 has the advantage over R_{pred}^2 because for very small samples, R_{pred}^2 is more prone to give negative values (although negative values of R_{resid}^2 are possible for discrete data). Therefore, for broad comparisons among different data sets using the same model, I would use R_{resid}^2 .

R^2 s are often used as “summary statistics” to describe the fit of a model to data in a way that does not involve statistical inference about the underlying stochastic process that generated the data: “How does the model fit these data?” rather than “How much does the model infer about the process that generated the data?” Should R^2 s be judged as a summary statistic? I think not. All the R^2 s showed high variation among simulations of the same model with the same parameters, especially when sample sizes were small (Fig. 4). This means that how the model fits a specific data set involves a lot of chance, and hence one should not get too excited about a high R^2 , or too discouraged about a low one. R^2 s are best treated as inferential statistics, that is, as functions of a data-generating process that are themselves random variables (Cameron and Windmeijer, 1996; Nakagawa and Schielzeth, 2013). As an inferential statistic, R_{lik}^2 most directly ties to hypothesis testing between full

and reduced models using a likelihood ratio test, and therefore it is particularly valuable as a partial R^2 . I think this makes R_{lik}^2 the best choice for investigating how well a specific model fits a specific data set using partial R^2 s.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.345v6>.

ACKNOWLEDGMENTS

I thank Ted Garland, Daijiang Li, Shinichi Nakagawa, Lucas Nell, Eric Pedersen, Joe Phillips, Bo Wang, Gang Wang, and especially Pierre de Villemereuil for wonderfully insightful comments that helped to clarify this article.

FUNDING

This work was supported by the US National Science Foundation [DEB-LTREB-1052160 and DEB-1240804].

APPENDIX

SCALING VARIANCES IN GLMMs AND PGLMMs

GLMMs and PGLMMs are formulated without a residual variance. Nonetheless, in algorithms used to fit GLMMs and PGLMMs, terms arise that can be interpreted as residual errors. These are important for the formulation of R^2 s in two ways. First, they explain why PGLMMs do not suffer the complications that arise for PGLS in defining R^2 s. The “residual errors” of PGLMMs never disappear, in contrast to the non-phylogenetic residuals in PGLS models that often go to zero. Second, the “residual errors” of GLMMs and PGLMMs provide a way of computing the scaling variance σ_d^2 in R_{resid}^2 (equation 11).

Residual Errors in Fitting GLMMs and PGLMMs

For simplicity, consider the case of a binary GLMM with only a single random effect and no fixed effects other than an intercept:

$$\begin{aligned} P\{y_i = 1 | b_i\} &= \mu_i \\ g(\mu_i) &= \beta + b_i \\ b_i &\sim \text{Gaussian}(0, \sigma_b^2) \end{aligned} \quad (\text{A1})$$

In GLMM fitting algorithms including those of Schall (1991), Breslow and Clayton (1993), and the Laplace algorithm used in `glmer()` (Pinheiro and Chao, 2006), the model is fit by constructing a covariance matrix in the Gaussian space (second line in equation A1) given

by $\mathbf{V} = \mathbf{W}^{-1} + \sigma_b^2 \mathbf{C}$. Here, $\sigma_b^2 \mathbf{C}$ is the covariance matrix for the random effects design matrix. \mathbf{W}^{-1} is the diagonal matrix containing the theoretical variances of $\varepsilon_i g'(\mu_i)$ where $\varepsilon_i = (y_i - \mu_i)$ is the residual in the data space and $g'(\mu_i)$ is the derivative of the link function evaluated at the conditional value μ_i corresponding to datum y_i . The matrix \mathbf{W} also appears in algorithms used to estimate GLMs, and its diagonal elements are known as the GLM iterated weights (McCullagh and Nelder, 1989). Because the $\varepsilon_i g'(\mu_i)$ s never all go to zero in a well-defined model, they are treated as “residual errors” in the Gaussian space.

The matrix \mathbf{V} can be derived formally from the likelihood function of the GLMM (Pinheiro and Chao, 2006). Heuristically, fitting the GLMM is analogous to fitting the Gaussian-space model

$$z_i = \beta + b_i + \varepsilon_i g'(\mu_i) \tag{A2}$$

The algorithms of Schall (1991) and Breslow and Clayton (1993) implement this expression iteratively, using the “working” estimates of μ_i to compute the residuals $\varepsilon_i g'(\mu_i)$, and then treating equation (A2) as a GLS regression to estimate $\beta + b_i$ and from these produce “updated” estimates of μ_i . It is necessary to do this iteratively, because in general the link function $g(\mu_i)$ is nonlinear, so the residuals $\varepsilon_i g'(\mu_i)$ depend on μ_i . The same approach was adapted for phylogenetic models by Ives and Helmus (2011). Because μ_i are generally different for each datum i , the variances of $\varepsilon_i g'(\mu_i)$, denoted $\sigma_{w[i]}^2$, will also differ for each i . This is depicted in Figure 1 as variable lengths of $\sigma_{w[i]}^2$ added to the tips of the covariance trees for the GLMM and PGLMM.

GLMM and PGLMM Scaling Variances σ_d^2 for R^2_{resid}

R^2_{resid} requires a scaling term σ_d^2 for the variances of fixed and random effects (equation 11). For binary GLMMs with a logit link function, Nakagawa and

Schielzeth (2013) use $\sigma_d^2 = \pi^2/3$, the variance of the logistic distribution, while they use $\sigma_d^2 = 1$ for the probit link function. Nakagawa et al. (2017) give additional approaches based on different approximation methods. Comparing estimates of the random effects variances for a binary GLMM fit with logit vs. probit link functions is a useful way to assess the scaling term σ_d^2 . Below I illustrate that neither $\sigma_d^2 = \pi^2/3$ nor the delta approximation from Nakagawa et al. (2017) give particularly good scaling terms for binary GLMMs with a logit link function. I also present two better alternatives.

To explore the accuracy of $\sigma_{d.NS}^2 = \pi^2/3$, consider the estimates of the random effects variance for equation (A1) using a logit vs. a probit link function, $\sigma_b^2[\text{logit}]$ and $\sigma_b^2[\text{probit}]$. If the scaling $\sigma_{d.NS}^2$ is correct, then $\sigma_b^2[\text{logit}]/\sigma_{d.NS}^2$ should approximate $\sigma_b^2[\text{probit}]$; they will not in general be exactly the same due to differences in the fits of the logit and probit GLMMs. In a simple simulation study (see Supplementary Section 2 available on Dryad), using $\sigma_{d.NS}^2 = \pi^2/3$ gives lower values $\sigma_b^2[\text{logit}]/\sigma_{d.NS}^2$ than $\sigma_b^2[\text{probit}]$, implying that $\sigma_{d.NS}^2 = \pi^2/3$ is too large (Fig. A1a).

There are two better alternatives. The first is $0.8768809 \pi^2/3$; call this the “reduced Nakagawa and Schielzeth” σ_d^2 denoted $\sigma_{d.rNS}^2$. $\sigma_{d.rNS}^2$ is better than $\sigma_{d.NS}^2$ (Fig. A1b) for reasons explained in the next section.

The second alternative comes naturally from the residual variances $\sigma_{w[i]}^2$ arising from the estimation algorithms for GLMMs. There is a technical complication: because $\sigma_{w[i]}^2$ take on different values for each datum i , how should these be combined to give a single σ_w^2 ? The algorithms for fitting GLMMs scale the total covariance matrix \mathbf{V} according to the log of its determinant, and this implies that σ_w^2 should be the geometric mean of the values of $\sigma_{w[i]}^2$. The

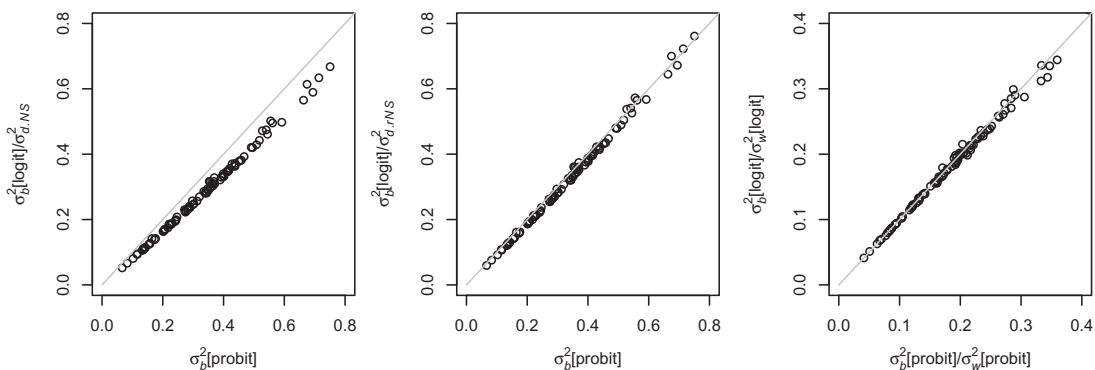


FIGURE A1. Accuracy of three scaling methods for the random effects variance in the binary GLMM in equation (A1). In the left panel, $\sigma_b^2[\text{logit}]/\sigma_{d.NS}^2$ is plotted against $\sigma_b^2[\text{probit}]$, where $\sigma_{d.NS}^2 = \pi^2/3$. In the center panel, $\sigma_b^2[\text{logit}]/\sigma_{d.rNS}^2$ is plotted against $\sigma_b^2[\text{probit}]$, where $\sigma_{d.rNS}^2 = 0.8768809\pi^2/3$. In the right panel, $\sigma_b^2[\text{logit}]/\sigma_w^2[\text{logit}]$ is plotted against $\sigma_b^2[\text{probit}]/\sigma_w^2[\text{probit}]$, where $\sigma_w^2[\text{logit}]$ and $\sigma_w^2[\text{probit}]$ are given in equation (A3). One-hundred data sets with 10 levels of the random effect and a total sample size of 1000 were simulated and fit using the code provided in Supplementary Section 2 available on Dryad.

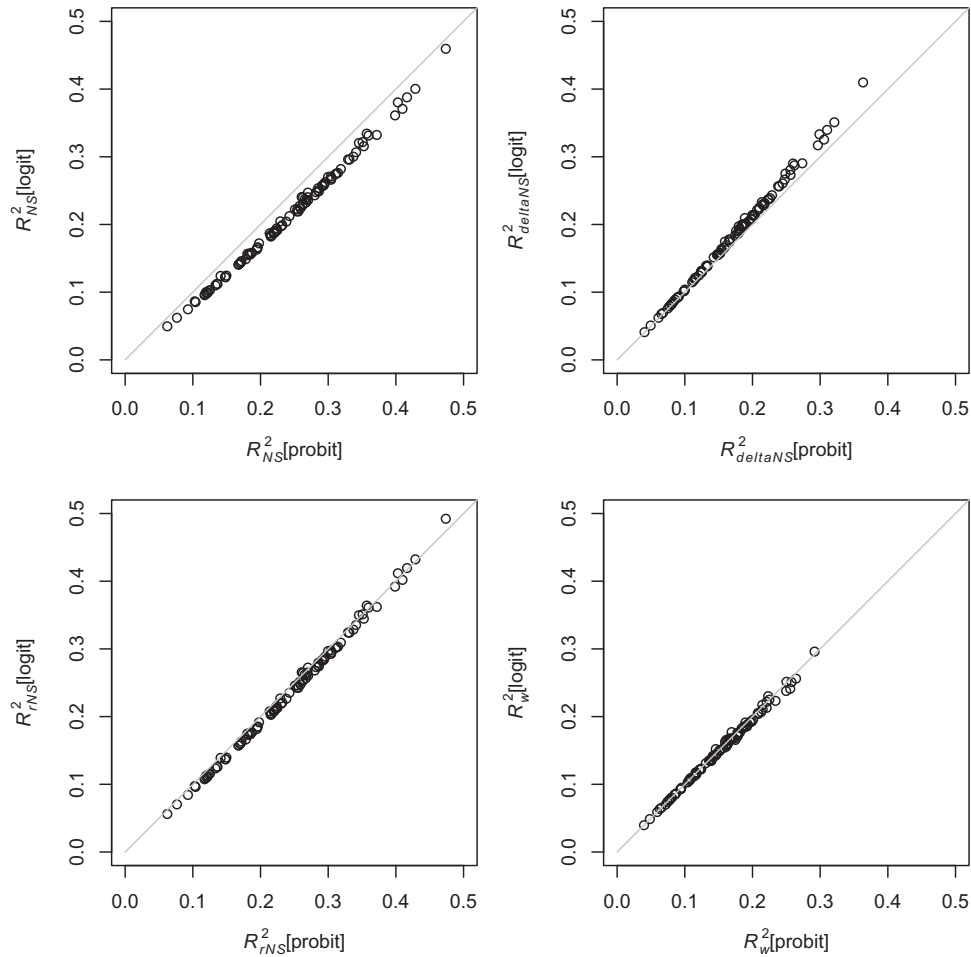


FIGURE A2. Accuracy of four scaling methods for R^2_{resid} in a GLMM (equation A1) determined by plotting values from a logit link function against values from a probit link function. One-hundred data sets with 10 levels of the random effect and a total sample size of 1000 were simulated and fit using the code provided in the [Supplementary Section 2](#) available on Dryad.

probit and logit random effects variances should be scaled by $\sigma_w^2[\text{probit}]$ and $\sigma_w^2[\text{logit}]$, respectively. Thus, $\sigma_b^2[\text{logit}]/\sigma_w^2[\text{logit}]$ should be approximately equal to $\sigma_b^2[\text{probit}]/\sigma_w^2[\text{probit}]$, which is the case (Fig. A1c).

From the performance of the scaling terms shown in Figure A1, R^2_{resid} computed with logit and probit link functions should be more similar using either $\sigma_{d,rNS}^2$ or σ_w^2 than using $\sigma_{d,NS}^2$. This is in fact the case (Fig. A2). Nakagawa et al. (2017) discuss a number of other approximations. In particular, although derived in a different manner, the values of σ_d^2 given by the delta method are similar to σ_w^2 , although they are computed using the estimates of the mean and variance of the distributional parameter (μ_i in the case of the binomial distribution), rather than the GLM iterated weights that differ among data points i . As shown in Figure A2, $\sigma_{d,rNS}^2$ and σ_w^2 outperform the delta-method values of σ_d^2 computed with `r.squaredGLMM()` in the MuMIn package (Barton, 2018). The performances of the methods for calculating

σ_d^2 depend on the details of the model, such as the sample size and the number of levels of the random effect; nonetheless, the patterns shown in Figure A2 are typical.

Using σ_w^2 as a scaling term produces lower values of R^2_{resid} than the other methods. This is a consequence of the other methods scaling $\sigma_d^2 = 1$ for the probit model, whereas σ_w^2 is less than 1 for probit models. Specifically, for binomial, Poisson, and GLMMs in general,

$$\begin{aligned} \sigma_w^2[\text{binomial}(\text{logit})] &= E_g\{n/(\mu_i(1-\mu_i))\} \\ \sigma_w^2[\text{binomial}(\text{probit})] &= E_g\{n\mu_i(1-\mu_i)/pdf(cdf(\mu_i))^2\} \\ \sigma_w^2[\text{poisson}] &= E_g\{1/\mu_i\} \\ \sigma_w^2[\text{general}] &= E_g\{g'(\mu_i)\text{var}(y_i-\mu_i)\} \end{aligned} \quad (\text{A3})$$

where $E_g\{\}$ is the geometric mean, and $pdf(\cdot)$ and $cdf(\cdot)$ are the probability density and cumulative density functions for a Gaussian distribution. An advantage of using σ_w^2 is that it generalizes to any distribution used in the

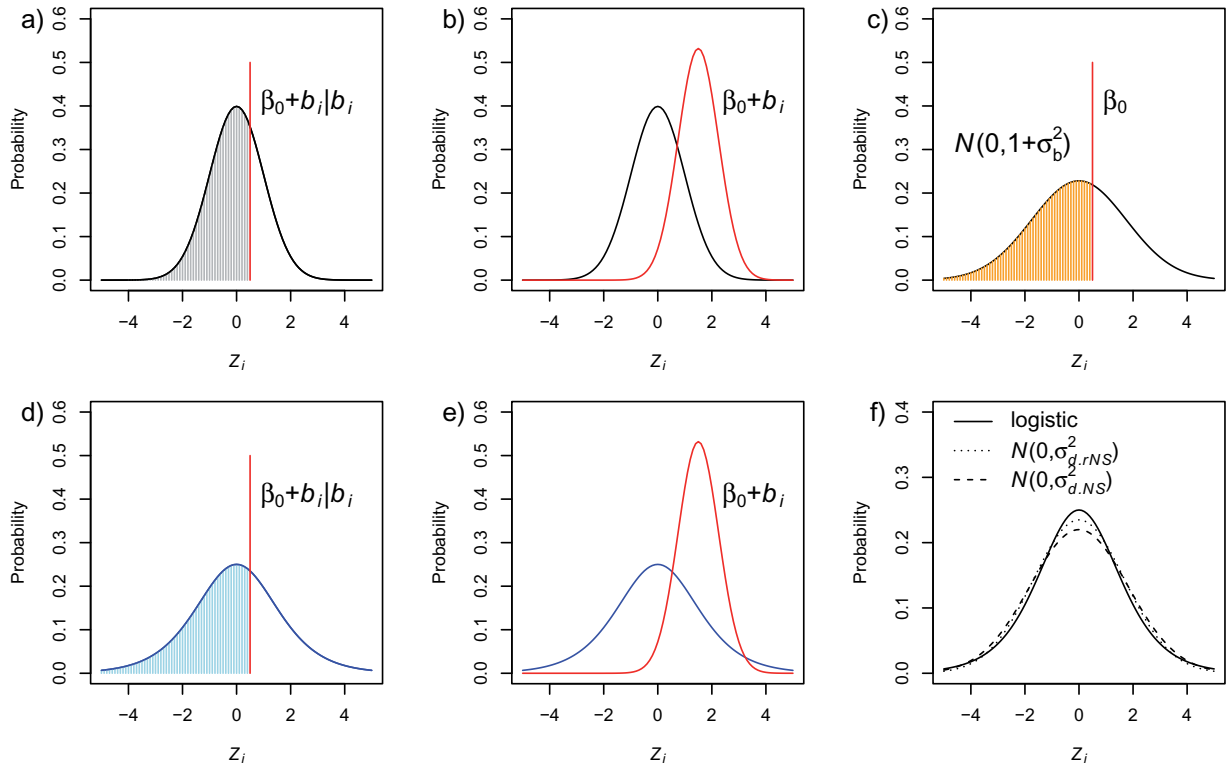


FIGURE A3. Threshold formulations of the GLMM in equation (A1) for probit (a–c) and logit (d–e) link functions. (f) Gives the probability density functions for a logistic distribution (solid), Gaussian distributions with variances $\sigma_{d,rNS}^2 = 0.8768809\pi^2/3$ dotted), and $\sigma_{d,NS}^2 = \pi^2/3$ dashed).

GLMM and is easy to compute from `glmer()` (Bates et al., 2015) and `binaryPGLMM()` (Ives and Garland, 2014). A disadvantage, however, is that it takes a different conceptual approach and gives lower values than $\sigma_{d,NS}^2 = \pi^2/3$ (Nakagawa and Schielzeth, 2013) that is becoming established in the literature. Furthermore, the value of σ_w^2 depends on the estimates of μ_i , and for small sample sizes or poorly fitting models, these estimates may not be well behaved.

The Threshold Formulation of GLMMs and $\sigma_{d,rNS}^2$

Where does $\sigma_{d,rNS}^2$ come from? The original $\sigma_{d,NS}^2 = \pi^2/3$ arises naturally when considering the threshold or latent-state formulation of GLMMs (for a historical summary, see Hadfield, 2015). The binary GLMM of equation (A1) can be equivalently formulated as

$$\begin{aligned}
 P\{Y_i = \mu_i | b_i\} &= F_D\{\beta + b_i | b_i\} \\
 b_i &\sim \text{Gaussian}(0, \sigma_b^2)
 \end{aligned}
 \tag{A4}$$

where F_D is the cumulative density function of the distribution corresponding to the link function; these are Gaussian and logistic distributions for probit and logit link functions, respectively. Equation (A4) can be seen as a threshold model by noting that $F_D\{\beta + b_i | b_i\}$ is the probability $P_D\{Z_i < \beta + b_i | b_i\}$ where $\beta + b_i | b_i$ serves as the threshold (e.g., Hadfield, 2015; de Villemereuil et al., 2016). This is depicted in Figure A3a for a probit link function in which $F_D\{Z_i = \beta + b_i | b_i\}$ is the shaded area

under Gaussian probability density function to the left of $\beta + b_i | b_i$. The threshold $\beta + b_i | b_i$ is conditional on the specific values of b_i . For the GLMM, the marginal value of the threshold $\beta + b_i$ is a Gaussian distribution with mean β and variance σ_b^2 (Fig. A3b). Therefore, the threshold model for a single point i can be written as a convolution

$$P\{Y_i = \mu_i\} = \int F_D\{Z_i | b_i\} p_N(b_i | \beta_0, \sigma_b^2) d\theta \tag{A5}$$

where $p_N(b_i | \beta_0, \sigma_b^2)$ is the Gaussian probability density function with mean β and variance σ_b^2 , and $Z_i = g(\mu_i)$. For the case of a probit link function when F_D is the cumulative density function of a $N(0, 1)$ distribution, this convolution is just the cumulative density function of a $N(\beta, 1 + \sigma_b^2)$ distribution, leading to another Gaussian threshold model (Fig. A3c).

For the logit link function, the same arguments follow (Fig. A3d,e). A complication occurs, however, because $F_D\{Z_i | b_i\}$ in equation (A5) is the cumulative density of a logistic distribution, rather than the Gaussian distribution as it is for the probit link function. The assumption $\sigma_{d,NS}^2 = \pi^2/3$ is akin to approximating the logistic $F_D\{Z_i | b_i\}$ with a Gaussian cumulative density function with the same variance, namely $\pi^2/3$. This is not very accurate, however, because the logistic distribution is leptokurtic relative to the Gaussian distribution (Fig. A3f). Instead, a better approximation can be made by selecting the variance of the Gaussian distribution to give a better match to the logistic cumulative density

function. A simple least-squares match gives $\sigma_{d,rNS}^2 = 0.8768809\pi^2/3$. While this is ad hoc, it works pretty well (Figs. A1 and A2).

Although this appendix has focused on GLMMs, identical arguments apply to PGLMMs.

REFERENCES

- Barton K. 2018. MuMIn: Multi-model inference R package version 1.42.1. <https://CRAN.R-project.org/package=MuMIn>.
- Bates D., Maechler M., Bolker B., Walker S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48.
- Bauwens D., Garland T., Castilla A.M., Vandamme R. 1995. Evolution of sprint speed in lacertid lizards—morphological, physiological, and behavioral covariation. *Evolution* 49:848–863.
- Blomberg S.P., Garland T. Jr., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Bolker B.M., Brooks M.E., Clark C.J., Geange S.W., Poulsen J.R., Stevens M.H.H., White J.S.S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24:127–135.
- Breslow N.E., Clayton D.G. 1993. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 88:9–25.
- Burnham K.T., Anderson D.R. 2002. Model selection and inference: a practical information-theoretic approach. 2nd ed. New York (NY): Springer.
- Buse A. 1973. Goodness of fit in generalized least-squares estimation. *Am. Stat.* 27:106–108.
- Cameron A.C., Windmeijer F.A.G. 1996. R-Squared measures for count data regression models with applications to health-care utilization. *J. Bus. Econ. Stat.* 14:209–220.
- Cameron A.C., Windmeijer F.A.G. 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* 77:329–342.
- Cox D.R., Snell E.J. 1989. The analysis of binary data. London, UK: Chapman and Hall.
- Cragg J.G., Uhler R.S. 1970. Demand for automobiles. *Can. J. Econ.* 3:386–406.
- de Villemereuil P., Schielzeth H., Nakagawa S., Morrissey M. 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics* 204:1281–1294.
- Edwards L.J., Muller K.E., Wolfinger R.D., Qaqish B.F., Schabenberger O. 2008. An R-2 statistic for fixed effects in the linear mixed model. *Stat. Med.* 27:6137–6157.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Garland T. Jr., Harvey P.H., Ives A.R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* 41:18–32.
- Gelman A., Hill J. 2007. Data analysis using regression and multi-level/hierarchical models. New York (NY): Cambridge University Press.
- Grafen A. 1989. The phylogenetic regression. *Trans. R. Soc. Lond. B Biol. Sci.* 326:119–157.
- Hadfield J.D. 2015. Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods Ecol. Evol.* 6:706–714.
- Ho L.S.T., Ane C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* 63:397–408.
- Housworth E.A., Martins E.P., Lynch M. 2004. The phylogenetic mixed model. *Am. Nat.* 163:84–96.
- Ives A.R. 2018. Mixed and phylogenetic models: a conceptual introduction to correlated data. Available from: <https://leanpub.com/correlateddata>. Accessed on August 14, 2018.
- Ives A.R., Garland T. 2010. Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* 59:9–26.
- Ives A.R., Garland T. Jr. 2014. Phylogenetic regression for binary dependent variables. In: Garamszegi L.Z. editor. Modern phylogenetic comparative methods and their application in evolutionary biology. Berlin Heidelberg: Springer-Verlag, p. 231–261.
- Ives A.R., Helmus M.R. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.* 81:511–525.
- Ives A.R., Li D. 2018. rr2 (R package). <https://github.com/arives/rr2>.
- 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.
- Jaeger B.C., Edwards L.J., Das K., Sen P.K. 2017. An R-2 statistic for fixed effects in the generalized linear mixed model. *J. Appl. Stat.* 44:1086–1105.
- Johnson P.C.D. 2014. Extension of Nakagawa & Schielzeth's R-GLMM(2) to random slopes models. *Methods Ecol. Evol.* 5:944–946.
- Judge G.G., Griffiths W.E., Hill R.C., Lutkepohl H., Lee T.-C. 1985. The theory and practice of econometrics. 2nd ed. New York: John Wiley and Sons.
- Kenward M.G., Roger J.H. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997.
- Kramer M. 2005. R2 statistics for mixed models. Manhattan (KS): New Prairie Press.
- Kvalseth T.O. 1985. Cautionary note about R2. *Am. Stat.* 39:279–285.
- Lavin S.R., Karasov W.H., Ives A.R., Middleton K.M., Garland T. Jr. 2008. Morphometrics of the avian small intestine, compared with non-flying mammals: a phylogenetic approach. *Physiol. Biochem. Zool.* 81:526–550.
- Liu H.H., Zheng Y., Shen J. 2008. Goodness-of-fit measures of R(2) for repeated measures mixed effect models. *J. Appl. Stat.* 35:1081–1092.
- Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45:1065–1080.
- Maddala G.S. 1983. Limited-dependent and qualitative variables in econometrics. Cambridge, UK: Cambridge University Press.
- Magee L. 1990. R2 measures based on Wald and likelihood ratio joint significance tests. *Am. Stat.* 44:250–253.
- 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.
- McCullagh P., Nelder J.A. 1989. Generalized linear models. 2nd ed. London: Chapman and Hall.
- Menard S. 2000. Coefficients of determination for multiple logistic regression analysis. *Am. Stat.* 54:17–24.
- Nagelkerke N.J.D. 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78:691–692.
- Nakagawa S., Johnson P.C.D., Schielzeth H. 2017. The coefficient of determination R-2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* 14:20170213.
- Nakagawa S., Schielzeth H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4:133–142.
- Orelien J.G., Edwards L.J. 2008. Fixed-effect variable selection in linear mixed models using R-2 statistics. *Comput. Stat. Data Anal.* 52:1896–1907.
- Pagel M. 1997. Inferring evolutionary processes from phylogenies. *Zool. Scr.* 26:331–348.
- Paradis E. 2012. Analysis of Phylogenetics and Evolution with R. New York: Springer.
- Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Version 4.1. *Bioinformatics* 20:289–290.
- Petersen K.B., Pedersen M.S. 2012. The matrix cookbook. Available from: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf. Accessed on Jan 27, 2015.
- Pinheiro J.C., Chao E.C. 2006. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *J. Comput. Graph. Stat.* 15:58–81.
- Revell L.J., Harmon L.J., Collar D.C. 2008. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* 57:591–601.
- Schall R. 1991. Estimation in generalized linear models with random effects. *Biometrika* 78:719–727.
- Xu R.H. 2003. Measuring explained variation in linear mixed effects models. *Stat. Med.* 22:3527–3541.