Running Head: ***R*^{2}s for Correlated Data**

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

Anthony R. Ives, Department of Integrative Biology, UW-Madison, Madison, WI 53706

arives@wisc.edu

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 variance is explained by phylogeny?" Here, I investigate three R^2s for phylogenetic and mixed models. A least-squares R^2_{ls} 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_{glmm} proposed by Nakagawa & Schielzeth (2013). The conditional expectation R^2_{ce} is based on "predicting" each residual from the remaining residuals of the fitted model. The likelihood ratio R^2_{lr} was first used by Cragg & Uhler (1970) for logistic regression, and here is used with the standardization proposed by Nagelkerke (1991). These three R^2s are formulated as partial R^2s, making it possible to compare the contributions of mean components (regression coefficients in phylogenetic models and fixed effects in mixed models) and variance components (phylogenetic correlations and random effects) to the fit of models. The properties

of the R^2s for phylogenetic models were assessed using simulations for continuous and binary response data (phylogenetic generalized least squares and phylogenetic logistic regression). Because the R^2s are designed broadly for any model for correlated data, the R^2s were also compared for LMMs and GLMMs. R^2_{ls}, R^2_{ce}, and R^2_{lr} all have good performance, and each has advantages and disadvantages for different applications. These R^2s are computed in the R package rr2 (https://github.com/arives/rr2). [Binomial regression, coefficient of determination, non-independent residuals, phylogenetic model, pseudo-likelihood]

INTRODUCTION

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 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. 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. The situation is more complex for models for discrete response variables, such as phylogenetic logistic regression models (PLOG) and generalized linear mixed models (GLMMs). For models of discrete distributions, even perfectly fitting models have residual variation due to the discreteness of the data, and this complicates the interpretation of an R^2.

Correlated errors in statistical models 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 (covariances) in the errors, but the phylogenetic signal does not directly lead to predictions of the fitted data. Desdevises *et al.* (2003) propose an R^2 in which a phylogeny is decomposed into principle components, and the principle components are used as predictor variables in a set of regressions; however, it is not clear how the resulting R^2 maps back to the statistical fit of a model and hence the statistical confidence in its results. An R^2 should ideally be tightly coupled to the fitted model and use this fit to quantify how much of the data are "explained" by the phylogeny.

Here, I assess three R^2s for models that specify non-zero covariances among errors. Although the definitions of R^2s 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 (PLOG) data. In addition, I will compare the properties of the R^2s applied to phylogenetic

and mixed models, which makes it possible to explore the R^2s in detail. This comparison also validates the R^2s as viable measures of goodness-of-fit for to a broad class of models. This is important, because R^2s should make it possible to assess and compare as wide a range of models as possible (Kvalseth 1985).

The general form of the investigated models is

$$Y_i \sim \mathcal{F}(\mu_i)$$

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

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma^2\mathbf{\Sigma}(\theta)) \qquad (1)$$

where data Y_i (i = 1, ..., n) are distributed by a member $\mathcal{F}$ of the exponential family of distributions (McCullagh & Nelder 1989). The parameter μ_i of distribution $\mathcal{F}$ 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 an error term e_i. The error term e_i has a multivariate Gaussian distribution with means 0 and covariance matrix $\sigma^2\mathbf{\Sigma}(\theta)$ that may depend on a parameter θ. When the link function $g()$ is the identify function, then equation (1) becomes a linear model (e.g., PGLS or LMM). PLOG can be modeled as a phylogenetic GLMM (PGLMM) in which Y_i has outcomes 0 and 1, and the link function $g()$ is logit (Ives & Helmus 2011; Ives & Garland 2014; Hadfield 2015). Note that other approaches to phylogenetic logistic regression (Ives & Garland 2010; Ho & Ane 2014) are not structured as PGLMMs, although calculating one of the three R^2s is still possible. Although equation (1) is written for only a single predictor variable x_i and parameter θ (a single random effect in a GLMM), all of the results presented below extend in the obvious way to multiple variables x_i and parameters θ.

The covariances in the residual errors are contained within $\sigma^2\mathbf{\Sigma}(\theta)$. In phylogenetic models, the structure of $\sigma^2\mathbf{\Sigma}(\theta)$ is typically generated under a specific model of evolution (Martins & Hansen 1997b; Lavin *et al.* 2008). For example, in a PGLS using Pagel's λ branch-length transform (Pagel 1997; Housworth, Martins & Lynch 2004), $\mathbf{\Sigma}(\lambda)$ is the sum of two matrices, $\mathbf{\Sigma}(\lambda) = \lambda\mathbf{\Sigma}_{\mathrm{BM}} + (1-\lambda)\mathbf{I}$, where $\mathbf{\Sigma}_{\mathrm{BM}}$ is the covariance matrix derived under the assumption of Brownian motion (BM) evolution (Felsenstein 1985; Grafen 1989), and $\mathbf{I}$ is the identity matrix. If λ=1, then the covariances between errors given by BM are proportional to the distance between two tips and their most common ancestor on the phylogenetic tree. If λ=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 adding tip branches to the BM phylogenetic tree that are scaled to make up a proportion λ of the base-to-tip distance (Fig. 1).

The similarity between PGLS and LMMs can be seen by depicting the LMM as a tree with branch lengths giving the strength of covariances among errors (Fig. 1). For a model with a single random effect b, the covariance matrix is $\mathbf{\Sigma}(\sigma^2_b) = \sigma^2_b\,\mathbf{\Sigma}_b + \mathbf{I}$ where $\mathbf{\Sigma}_b$ is a block-diagonal matrix whose values are 1 for each row i and column j corresponding to the same level (block) of the random effect (Gelman & Hill 2007). The greater the variance of the random effect σ^2_b, the greater the covariances among errors within the same level, and the smaller the relative contribution of the residual errors given by the length of the tips of the tree. For comparison with LMMs, GLMMs can also be depicted as a tree, but rather than the residual errors at the tips of the tree having length 1, for models of discrete data the error variance depends on the unavoidable differences between the observation (0 or 1 for binary data) and probability of the observation (taking any value between 0 and 1). The lengths of these variances σ^2_w depend on the probability of the observation. I am showing the variances σ^2_w only for illustrating the

113 similarities between LMM and GLMM models, and by extension PGLS and PLOG models. In
114 some methods for implementing GLMMs (Schall 1991; Breslow & Clayton 1993), σ^2_w (the
115 inverses of the GLM weights; McCullagh & Nelder 1989) are used in the fitting algorithms; in
116 other methods, for example those in the R package 'lme4' (Bates *et al.* 2014), they are not used,
117 although they can nonetheless be extracted from fitted models.

118 The three R^2s presented here partition the "explained" and "unexplained" variances for
119 models with correlated errors such as those depicted in figure 1. Because models can contain
120 multiple parameters, the R^2s compare a full model with a reduced model in which one or more of
121 the parameters are removed; thus, they are partial R^2s that give the explained variance by the
122 components that differ between full and reduced models. The total R^2s are obtained by selecting
123 the reduced model in which there is only an intercept and residuals are independent. Defining
124 partial R^2s has the advantage of being able to ask about the contribution of a single or subset of
125 components to the fit of a model. This makes it possible to exclude coefficients in a model that
126 are not of explicit interest; for example, many phylogenetic models for species traits include
127 body size as one of the predictor variables to factor out body size, and partial R^2s make it
128 possible to assess the goodness-of-fit for the remaining predictor variables. By comparing a
129 model with a phylogeny to a model without, partial R^2s also make it possible to answer the
130 question "How much of the data is explained by phylogeny?"

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

136 for model selection (Burnham & Anderson 2002). Therefore, the LLR is a natural choice to
137 assess R^2s: a good R^2 should be monotonically related to the LLR. Second, can the R^2 separate
138 the contribution of different components of the model to the overall model fit? For the simple
139 case of equation (1) in which there is only a single regression coefficient (β_1) and a single
140 variance parameter (θ), I ask whether the R^2s can distinguish between the two in their
141 contributions to the fit of the model. Third, does the R^2 give similar values when applied to data
142 generated by the same statistical process? If the values of R^2 when applied to data generated from
143 the same statistical process are all similar, then the R^2 gives a precise measure of goodness-of-fit.
144

145 ## MATERIALS AND METHODS

146 There is an extensive literature on R^2s for GLMs and LMMs, and a growing literature for
147 GLMMs (Buse 1973; Cameron & Windmeijer 1996; Cameron & Windmeijer 1997; Kenward &
148 Roger 1997; Menard 2000; Xu 2003; Kramer 2005; Edwards *et al.* 2008; Liu, Zheng & Shen
149 2008; Orelien & Edwards 2008; Nakagawa & Schielzeth 2013; Jaeger *et al.* 2017), and this
150 literature forms the basis for the R^2s that can be applied to phylogenetic models. The three R^2s
151 take three different approaches to defining "explained variance", the same general approaches
152 considered for LMMs by Xu (2003). R^2_{ls} is based on the variance of the residuals in a way that
153 explicitly incorporates their covariances. For models with discrete data, R^2_{ls} is defined to closely
154 match R^2_{glmm} presented by Nakagawa & Schielzeth (2013). R^2_{ce} is based on the difference
155 between the observed data and model predictions (Kvalseth 1985), where the predictions use
156 information from the covariances among errors. R^2_{lr} is based upon the information that is gained
157 by adding parameters (regression parameters or covariance terms); it uses the likelihood ratio of
158 full to reduced models as was first proposed for logistic regression (Cragg & Uhler 1970;

159 Maddala 1983; Cox & Snell 1989). For ordinary linear models without correlated errors, these

160 three R^2s are identical to the OLS R^2, but they differ for models with correlated errors.

161 Here I give a heuristic explanation for the R^2s, and Appendix 1 gives details about their

162 implementation. The starting point in the derivations of the three R^2s is the standard R^2 for

163 continuous data (Buse 1973; Judge *et al.* 1985),

164 $$R^2 = 1 - \frac{\mathrm{mSE}_f\left(\hat{\theta}_f\right)}{\mathrm{mSE}_r\left(\hat{\theta}_r\right)} \tag{2}$$

165 where mSE_f is the mean squared errors for the full model, and mSE_r is for the reduced model.

166 For the unadjusted R^2, the mSEs are the mean SEs without correcting for degrees of freedom, so

167 I have used the abbreviation mSE rather than the normal MSE, the mean squared errors corrected

168 for degrees of freedom. Both full and reduced models may contain parameters in vectors

169 θ_f and θ_r that involve the variances and covariances among samples which are estimated when

170 the model is fit; for calculating the mSE, these parameters are assumed to be fixed. For a

171 generalized least-squares model (Judge *et al.* 1985),

172 $$\mathrm{mSE}\left(\hat{\theta}\right) = \frac{1}{n}\left(\mathbf{Y} - \mathbf{X}\hat{\beta}\right)' \mathbf{V}\left(\hat{\theta}\right)^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\beta}\right) \tag{3}$$

173 where $\mathbf{Y}$ is the $n \times 1$ vector of response values Y_i, $\mathbf{X}$ is the $n \times p$ matrix for p predictor variables

174 (including the intercept), $\hat{\beta}$ is the $1 \times p$ vector of estimated regression coefficients (fixed effects),

175 and $\mathbf{V}\left(\hat{\theta}\right)^{-1}$ is the inverse of the $n \times n$ matrix $\mathbf{V}(\theta) = \sigma^2\mathbf{\Sigma}(\theta)$ that contains the variances and

176 covariances of the errors. The mSE for OLS models is the special case in which $\mathbf{\Sigma}(\theta) = \mathbf{I}$, which

177 gives the standard R^2.

178 The key issue in defining R^2s is the scaling of $\mathbf{V}(\theta) = \sigma^2\mathbf{\Sigma}(\theta)$. Setting $\mathbf{V}(\theta) = \mathbf{\Sigma}(\theta)$ in

179 equation (3), the mSE gives the maximum likelihood estimate of the variance term σ^2 from

equation (1). However, $\mathbf{\Sigma}(\theta)$ can be rescaled by a constant without changing the fit of the statistical model; the only effect of multiplying $\mathbf{\Sigma}(\theta)$ by a constant is to change the value of σ^2 by 1/constant. This would not be an issue if the scaling were the same for full and reduced models, because the scaling would cancel out when dividing mSE_f by mSE_r. However, it will generally be the case that $\mathbf{\Sigma}(\hat{\theta}_f) \neq \mathbf{\Sigma}(\hat{\theta}_r)$; for example, even for LMMs that include the same random effects, $\mathbf{\Sigma}(\hat{\theta}_f) \neq \mathbf{\Sigma}(\hat{\theta}_r)$ if removing fixed effects from the full model changes the estimated variances of the random effects in the reduce model. Therefore, the scaling is not removed by dividing mSE_f by mSE_r. Because the scaling determines the estimate of σ^2, it affects the R^2. The three R^2s presented below differ in how they address the issue of scaling $\mathbf{\Sigma}(\theta)$.

R^2_{ls}

R^2_{ls} (for least squares) extends equation (2) in a way that closely matches R^2s that have been proposed for LMMs and GLMMs (Buse 1973; Xu 2003; Edwards *et al.* 2008; Nakagawa & Schielzeth 2013; Jaeger *et al.* 2017). For LMMs, a natural scaling of $\mathbf{\Sigma}(\theta)$ is to let $\mathbf{\Sigma}(\theta) = \mathbf{I} + \mathbf{G}(\theta)$ where $\mathbf{G}(\theta)$ is the block-diagonal matrix containing the variances of the random effects divided by the residual variance (i.e., σ^2_b in Fig. 1). With this scaling, the residual variance σ^2 estimated from the LMM equals exactly mSE from equation (3). For the total R^2, the reduced model contains only the intercept and therefore equals the variance of the data. This means that the total R^2_{ls} for LMMs differs slightly from $R^2_{glmm(c)}$ (Nakagawa & Schielzeth 2013) in which the variance in the data is estimated from the full model, rather than the actual variance of the data. Nonetheless, the values of R^2_{ls} and $R^2_{glmm(c)}$ will be very close.

For phylogenetic models, scaling $\mathbf{\Sigma}(\theta)$ is more complicated. For some common branch-length transforms in PGLS that are used to measure the strength of phylogenetic signal, such as the OU transform (Martins & Hansen 1997a; Blomberg, Garland & Ives 2003), the covariance matrix $\mathbf{\Sigma}(\theta)$ does not separate additively to give terms for the explained versus unexplained variance. Even though Pagel's λ branch-length transform, $\mathbf{\Sigma}(\lambda) = \lambda\mathbf{\Sigma}_{BM} + (1-\lambda)\mathbf{I}$, does break down into the sum of phylogenetic and non-phylogenetic terms, the non-phylogenetic term $(1-\lambda)\mathbf{I}$ cannot be interpreted as the unexplained variance. This is because, for many data sets, the estimate of λ will be 1, which is the expectation under the assumption of BM evolution. If this occurs, this would force an R^2 treating $(1-\lambda)\mathbf{I}$ as the unexplained variance to be 1, regardless of the explanatory power of the predictor variances (fixed effects); this property would make the R^2 uninformative. To solve this problem, I propose scaling $\mathbf{\Sigma}(\theta)$ so that the total branch lengths equal 1; this is equivalent to assuming that the total amount of independent evolution is the same. For a fitted tree with strong phylogenetic signal, scaling $\mathbf{\Sigma}(\theta)$ to have the total 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 $\mathbf{\Sigma}(\theta)$ for 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_{ls} performs for PGLS models in a similar way as it does for LMMs.

For discrete models, it is necessary to account for the variation introduced by discrete data (Fig. 1). While there are different ways to do this (Appendix 1), an approach that makes R^2_{ls} conform to $R^2_{glmm(c)}$ for GLMMs (Nakagawa & Schielzeth 2013) is to replace mSE in equation (2) with $\frac{\sigma_d^2}{\hat{\sigma}_f^2 + \hat{\sigma}_l^2 + \sigma_d^2}$ where σ_d^2 gives the distribution-specific variance attributed to the

223 discreteness of the data; $\sigma_d^2 > 0$ because the predictions from the model are probabilities whereas

224 the observations are discrete values (0 or 1 for binary data). For binary data $\sigma_d^2 = \pi^2/3$. The

225 variance that σ_d^2 measures is at the level of the transformed parameter value (i.e., $g(\mu_i)$ in

226 equation (1)) for which the errors are normally distributed. To scale the variance σ_d^2 by the

227 variances of the data in the transformed space of $g(\mu_i)$, it is necessary to divide σ_d^2 by the total

228 variance that is given by the regression coefficients (fixed effects), $\hat{\sigma}_f^2$, and the errors, $\hat{\sigma}_l^2$. For

229 GLMMs $\hat{\sigma}_l^2$ is the estimate of the variance of the random effects. For PGLMMs $\hat{\sigma}_l^2$ is the

230 estimate of the variance in the Gaussian error term $\sigma^2\mathbf{\Sigma}$ in which $\mathbf{\Sigma}$ is scaled to have base-to-tip

231 branch lengths of 1. Note that in contrast to $R^2_{glmm(c)}$ (Nakagawa & Schielzeth 2013), R^2_{ls} is

232 explicitly defined as a partial R^2; this not only makes it possible to assess the contributions of

233 different components of the model to goodness-of-fit, it also simplifies application and

234 interpretation of R^2_{ls} to LMMs and GLMMs with complex random effects, such as random slope

235 effects (Johnson 2014).

236

237 R^2_{ce}

238 R^2_{ce} (for conditional expectation) is based on the variance in the difference between

239 observed and predicted data, $\frac{1}{n}\sum\left(Y_i - \hat{Y}_i\right)^2$. This approach conceptually comes the closest to

240 answer the question of how much variation in the data is explained by the covariances in the

241 model. For the case of LMMs, the predicted values $\hat{Y}_i$ can be taken as the sum of the fixed

242 effects and the estimate of the value of the random effects. For the LMM in figure 1, this

243 corresponds to the estimated value at the polytomy formed at the node shared by all observations

244 within the same level of the random effect. As the number of observations within each level of
245 the random effect increases, R^2_{ce} for LMMs converges to R^2_{ls} because the estimates of the values
246 of the random effects become more precise.

247 As it does for R^2_{ls}, PGLS poses a complication for R^2_{ce}: what is the predicted value $\hat{Y}_i$?
248 To parallel R^2_{ce} for LMMs, $\hat{Y}_i$ could be taken as the estimated value at the node immediate below
249 the tip on the phylogeny containing Y_i. For phylogenies with some short terminal branch lengths,
250 however, the estimates for the node underneath Y_i will be determined largely by the value of Y_i
251 itself, leading to very high (and uninformative) R^2s. Therefore, for PGLS I defined R^2_{ce} using the
252 estimates $\hat{Y}_i$ computed by removing the point Y_i from the data set and then estimating $\hat{Y}_i$ from
253 the predictor variables and the remaining data points. Specifically for equation (1), the expected
254 value of residual $R_i = Y_i - (\beta_0 + \beta_1 x_i)$ from the remaining residuals $\mathbf{R}_{[-i]}$ is

255 $$\hat{R}_i = \bar{R} + \mathbf{V}_{[i,-i]}\mathbf{V}^{-1}_{[-i,-i]}\left(\mathbf{R}_{[-i]} - \bar{R}\right), \quad (4)$$

256 where $\bar{R}$ is the GLS mean of the residuals, $\mathbf{V}_{[i,-i]}$ is row i of $\mathbf{V}$ with column i removed, and $\mathbf{V}_{[-i,-i]}$
257 is $\mathbf{V}$ with row i and column i removed (Petersen & Pedersen 2012). The predicted value of Y_i is
258 then $\hat{Y}_i = \beta_0 + \beta_1 x_i + \hat{R}_i$. Note that this procedure (removing Y_i before predicting $\hat{Y}_i$) could be
259 used for LMMs (and other models), although to make R^2_{ce} conform most closely to the structure
260 of LMMs, the LMM R^2_{ce} makes predictions $\hat{Y}_i$ while keeping Y_i in.

261 The same approach as used for LMMs can by used for GLMMs and PGLMMs. In these
262 cases, the variances are calculated for untransformed values of Y_i, rather than in the Gaussian
263 space of $g(\mu_i)$ as was done for R^2_{ls}. These predicted values are given from the estimation
264 algorithms for GLMMs by glmer in the R package lme4 (Bates *et al.* 2014) and for PGLMMs by
265 binaryPGLMM (Ives & Helmus 2011; Ives & Garland 2014) in the R package ape (Paradis &

Paradis 2012). Because GLMMs and PGLMMs include the unavoidable variance in $Y_i - \hat{Y}_i$ due to the discreteness of the data, the estimates $\hat{Y}_i$ correspond to those values a distance σ^2_w from the tips of the tree in figure 1.

R^2_{lr}

R^2_{lr} (for likelihood ratio) is the application of an R^2 proposed for logistic regression (Cragg & Uhler 1970; Maddala 1983; Cox & Snell 1989) and generalized by Magee (1990) and Nagelkerke (1991), and used for LMMs by Kramer (2005). R^2_{lr} computed for a range of models in the MuMIn package of R (Barton 2016). For LMMs, R^2_{lr} differs from R^2_{ls} only in the scaling of $\mathbf{V}(\theta)$. If $\mathbf{V}(\theta)$ is scaled so that the determinant $\det(\mathbf{V}(\theta)) = 1$, then the maximum log likelihood is

$$logLik\left(\hat{\theta}\right) = -\frac{n}{2}\left(\log\left(2\pi \text{ mSE}\left(\hat{\theta}\right)\right)+1\right). \tag{5}$$

Substituting into equation (2) then leads to

$$R^2_{lr} = 1 - \exp\left(\frac{-2}{n}\left(logLik\left(\hat{\theta}_f\right) - logLik\left(\hat{\theta}_r\right)\right)\right). \tag{6}$$

This definition of R^2_{lr} in terms of likelihoods extends immediately to any model fit by maximum likelihood estimation, including PGLS and GLMM. However, for discrete data equation (6) does not have a maximum of 1, because the maximum attainable log-likelihood for discrete data is zero. Therefore, Nagelkerke (1991) and Cameron and Windmeijer (1997) proposed dividing by the maximum attainable value, which is equation (6) with $logLik\left(\hat{\theta}_f\right) = 0$; throughout, I have used this Nagelkerke standardization.

286 The algorithm used by binaryPGLMM to fit equation (1) for binary phylogenetic data
287 uses quasi-likelihoods and does not give a true maximum likelihood that could be used to
288 compute R^2_{lr}. Therefore, for PLOG models I used phyloglm in the R package phyloglm (Ho &
289 Ane 2014), fitting the model with penalized ML but then using the provided ML values to
290 calculate R^2_{lr}.
291
292 *Simulations for assessment*
293 The simulations to assess the statistical properties of the R^2s applied to LMM, PGLS,
294 GLMM, and PLOG all follow the same strategy. For each, data were simulated using equation
295 (1) for the case with variation in a predictor variable and no covariances ($\beta_1 > 0$, $\theta = 0$), only
296 covariances ($\beta_1 = 0$, $\theta > 0$), or both ($\beta_1 > 0$, $\theta > 0$). For each case, the model parameters were the
297 same for all simulations, so that variation in values of a given R^2 among datasets is caused by
298 random sampling from the same statistical process.
299 For LMM, data were simulated with the model
300 $$y_i = \beta_0 + \beta_1 x_i + b u_i + e_i \quad (7)$$
301 where x_i follows a Gaussian distribution with mean 0 and variance 1, and the random effect u_i
302 has 10 levels, with b following a normal distribution with mean 0 and variance θ. I selected
303 parameter values to generate moderate R^2 values. For GLMMs, values from equation (7) without
304 the residual error term e_i were used through a logit link function (equation (1)) to produce
305 binomial probabilities for a binary model. Models were fit using lmer and glmer in the lme4
306 package of R (Bates *et al.* 2014).
307 For the PGLS model, to obtain the covariance matrix $\mathbf{\Sigma}(\theta)$ in equation (1), I first
308 simulated random phylogenetic trees using the rtree function of the ape package of R (Paradis,

Claude & Strimmer 2004). Thus, a different tree was simulated for each dataset. The strength of phylogenetic signal was varied using Pagel's λ transformation which served as the variance parameter θ. Values of x_i were simulated under the BM assumption using the rTraitCont function (Paradis, Claude & Strimmer 2004). The simulated data were fit using penalized maximum likelihood with the function phylolm assuming a Pagel's lambda transformation in the package phyloglm in R (Ho & Ane 2014). The PLOG model 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 datasets, making the simulation studies difficult. Phylogenetic signal in the residuals e_i was controlled by setting $\mathbf{\Sigma}(\theta) = \theta\mathbf{\Sigma}_{\mathrm{BM}}$ 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 in equation (1).

Simulated examples

I illustrate the three R^2s for two simulated examples. The first involves a PGLS model for two predictor variables and computes the partial R^2s for one of them. For example, suppose a researcher has data on sprint speed in lizards, and the predictor variables are hind leg length and body size (Bauwens *et al.* 1995). Hind leg length is the variable of interest, while body size is a covariate, 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 body size) either did or did not show phylogenetic signal. Specifically, the model was

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

$$e \sim \mathrm{Gaussian}(\mathbf{0}, \sigma^2\mathbf{\Sigma}_{\mathrm{BM}}) \qquad (8)$$

where log(*body size*) was selected from a Gaussian distribution with covariance matrix $\mathbf{\Sigma}_{BM}$, and log(*hind leg length*) was selected from a Gaussian distribution with covariance matrix either $\mathbf{I}$ or $\mathbf{\Sigma}_{BM}$. The data were fit using phylolm with Pagel's λ transformation.

As a second example, I simulated LMMs and binary GLMMs using equation (7), and fit them not only as LMM and GLMMs, but also as PGLS and PLOG models. The fitting as phylogenetic models was performed by converting the covariance matrix given by the random effects in the LMM and GLMM into a phylogenetic tree (Fig. 1) using the vcv2phylo function in ape (Paradis & Paradis 2012). This simulation allows a direct comparison between the R^2s applied to mixed versus phylogenetic models.

RESULTS

The R^2s were assessed according to the 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 partition sources of variation in the model as partial R^2s, and (iii) how precise is their inference about goodness-of-fit. Property (iii) treats the R^2s as if they were estimators of goodness-of-fit and asks how variable are the estimates when applied to repeated simulations from the same model (e.g., ; Cameron & Windmeijer 1996). A more comprehensive assessment is given in Appendix 2.

Goodness-of-fit

Figure 2 plots the total R^2s against the corresponding LLR. All R^2s were positively related to the LLR, which is a minimum requirement for an R^2. R^2_{lr} shows a monotonic relationship with LLR, which is necessarily the case due to the definition of R^2_{lr} (equations (5),

355 (6)). For the remaining R^2s, values for a given LLR were generally lower for simulations in
356 which variation was produced only by the fixed effect ($\beta_1 > 0$, $\theta = 0$; Fig. 1, blue circles). This
357 implies that, relative to the LLR, these R^2s were attributing less "explained" variance to
358 regression coefficients (fixed effects) than covariances parameters (phylogeny and random
359 effects).

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

364 All of the R^2s other than R^2_{lr} showed greater scatter in their relationships with LLR for
365 the simulations of binary data (GLMM and PLOG). In part, this is due to the difficulty of
366 estimating variance parameters θ in binomial models. For example, there is more scatter in R^2_{lr}
367 for GLMM simulations than LMM simulations, even though the criterion for calculating the R^2s
368 (the log likelihoods) are the same. The scatter seems particularly large for R^2_{ls} and R^2_{ce} applied to
369 PLOG simulations, although this case requires some technical discussion. For PLOG, the LLR
370 was obtained from phyloglm using penalized maximum likelihood, whereas R^2_{ls} and R^2_{ce} were
371 estimated from the model fit by binaryPGLMM using the pseudo-likelihood. The phyloglm
372 estimate of phylogenetic signal, λ, tended to absorb at zero even when the estimate λ from
373 binaryPGLMM was positive; therefore, R^2_{ls} and R^2_{ce} could be positive even when the LLR was
374 zero. Previous comparison between phyloglm and binaryPGLMM showed that they have similar
375 performances but do not necessarily give the same conclusions about the presence of
376 phylogenetic signal for the same dataset (Ives & Garland 2014).
377

Partitioning sources of variation

The partial R^2_{ls}, R^2_{ce}, and R^2_{lr} were generally able to partition sources of variation between components of a model, in particular between regression coefficients (fixed effects) and covariance parameters (random effects). Simulations with $\beta_1 > 0$ and $\theta = 0$ should have partial R^2s for β_1 that are positive and partial R^2s for θ that are zero (blue circles, Fig. 2). Simulations with $\beta_1 = 0$ and $\theta > 0$ should have partial R^2s for β_1 that are zero and partial R^2s for θ that are positive (red triangles, Fig. 2). Simulations with $\beta_1 > 0$ and $\theta > 0$ should have both partial R^2s > 0 (black x's, Fig. 2). Because the values of β_1 and θ were the same whether or not the other was zero, the partial R^2s for β_1 should be the same for simulations with $\theta = 0$ (blue circles) as for simulations with $\theta > 0$ (black x's), and the partial R^2s for θ should similarly be the same for $\beta_1 = 0$ (red triangles) and $\beta_1 > 0$. For continuous data (LMM and PGLS), all three R^2s had similar performance and similar values of the partial R^2s (see also Appendix 2). For binary data (GLMM and PLOG), the three R^2s 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 PLOG simulations with $\beta_1 > 0$ and $\theta > 0$ which sometimes gave a partial R^2_{lr} for θ of zero (black 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 underlying process

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

401 160 samples (16 replicates at each level) and then randomly removing two replicates at each
402 level to reduce the sample size in steps of 20. For PGLS and PLOG, each dataset at each sample
403 size was simulated independently.

404 For LMM simulations, R^2_{ls}, R^2_{ce}, and R^2_{adj} showed similar patterns (Fig. 4), reflecting the
405 fact that they give very similar values (Fig. 2, Appendix 2). Mean values did not change with
406 sample size, and there was only moderate increase in variability among simulations with
407 decreasing sample size. In contrast, mean values of R^2_{lr} decreased with decreasing sample size.
408 This probably reflects the information that is lost when estimating the model parameters. In
409 contrast to LMM simulations, the PGLS simulations showed less change in the means of R^2_{lr} and
410 R^2_{ce} with sample size, presumably because there were more covariances among samples (i.e., the
411 covariance matrix had more non-zero elements) than in the LMM with few replicates per level.

412 For the GLMM, both R^2_{ls} and R^2_{ce} had somewhat higher variances (less precision) than
413 R^2_{lr}. The greater variation in values of R^2_{ls} and R^2_{ce} compared to R^2_{lr} may occur because R^2_{ls} and
414 R^2_{ce} depend on estimates from the models (random effects for R^2_{ls} and fitted values for R^2_{ce})
415 whereas R^2_{lr} depends on likelihoods. Thus, R^2_{ls} and R^2_{ce} are compromised when the estimates are
416 poor, as is particularly the case when sample sizes are small. For PLOG, the variances in R^2_{ls} and
417 R^2_{ce} were similar to R^2_{lr} (Fig. 4). This is likely because estimates of phylogenetic signal ($\lambda = \theta$)
418 were well-bounded, in contrast to the variance in random effects in the GLMMs.

419

420 *Simulated examples*

421 In the simulated example of sprint speed regressed on log body mass (x_1) and log hind
422 limb length (x_2) (equation (8)), whether or not log hind limb length showed phylogenetic signal
423 had a large effect on the partial R^2s for the effect of log hind limb length (Table 1). In the fitted

PGLS models for both datasets, 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 ($P = 0.012$ when x_2 had phylogenetic signal and $P << 0.001$ when it did not). The partial R^2_{ls}, R^2_{ce}, and R^2_{lr} for hind leg length were 0.21-0.23 when hind leg length had phylogenetic signal, and 0.71-0.89 when it did not. In contrast, the partial R^2s for phylogenetic signal (reduced model with $\lambda = 0$) and the total R^2s (reduced model with $x_1 = x_2 = \lambda = 0$) did not differ much between simulations. The partial R^2s 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^2s that can assess the role of predictor variables separately from other variables like body size that are not of specific interest.

In the second example (Table 2), a LMM and a GLMM were simulated and then the data were fit using LMM and GLMM, and also PGLS and PLOG by converting the covariance matrix given by the random effects into a phylogeny (Fig. 1). R^2_{ls}, R^2_{lr}, and R^2_{ce} were computed for the total model, as well as partial R^2s for the fixed effect x and the random effect θ. As expected, R^2_{lr} was the same for mixed and phylogenetic models. All values for R^2_{ce} were also close between mixed and phylogenetic models. For the LMM simulation, R^2_{ls} calculated for the LMM model was very close to both R^2_{ce} and the R^2_{ols} computed by treating the random effect as a categorical fixed effect. However, the values of R^2_{ls} calculated from PGLS were lower. This occurred because the scaling of the covariance matrix $\mathbf{\Sigma}$ for LMM and PGLS is different; for LMM R^2_{ls} scales $\mathbf{\Sigma}$ so that the residual error corresponds to 1, whereas for PGLS R^2_{ls} scales $\mathbf{\Sigma}$ so that the total branch lengths equal 1. Although this gives different values of R^2_{ls} for LMM and PGLS, it

447 avoids forcing R^2_{ls} to equal 1 when the residual error matches its expectation under BM
448 evolution. For the GLMM simulation, R^2_{ls} from the fitted GLMM is higher than R^2_{ls} from the
449 fitted PLOG, although R^2_{ls} for both models are higher than R^2_{lr} and R^2_{ce}.
450

451 ## DISCUSSION

452 R^2_{ls}, $R^2_{ce,}$ and R^2_{lr} are presented here with focus on phylogenetic models, although they
453 are broadly applicable to models with correlated errors. Below, I first address their specific
454 application to phylogenetic models using mixed models as a reference, and then give general
455 recommendations.
456

457 ### *Applications to LMM, PGLS, GLMM, and PLOG*

458 For both continuous data (LMM and PGLS) and discrete data (GLMM and PLOG), all
459 R^2s had good performance. For the simple model with a single regression coefficient β_1 and
460 single covariance parameter θ (equation (7)), all R^2s were reasonable measures of goodness-of-fit,
461 as assessed against the log likelihood ratio between full and reduced models (Fig. 2).
462 Nonetheless, R^2_{ls} and R^2_{ce} gave lower partial R^2 values for the regression coefficient β_1 relative to
463 the partial R^2 values for the covariance parameter θ in comparison to R^2_{lr} and the log likelihood
464 ratio, LLR (Fig. 2). Also, although all three R^2s gave very similar values for the same dataset
465 with continuous data, the values differed more for discrete data fit with either GLMM or PLOG
466 (Fig. 2 and Appendix 2). This is reflected in general by the decreased precision of the R^2s applied
467 to discrete data, as measured by the variation in values when fit to data simulated under the same
468 parameter values (Fig. 4). All R^2s were capable of identifying whether β_1 or θ was responsible
469 for the fit of the model to the data as determined by the partial R^2s; when β_1 or θ were zero in the

470 simulations, the partial R^2s for β_1 or θ, respectively, were low (Fig. 3). However, the partial R^2s
471 for GLMM and PLOG tended to be more variable and less conclusive than for LMM and PGLS
472 (Fig. 3). Finally, R^2_{lr} decreased as sample sizes decreased especially for LMMs but also for
473 GLMMs (Fig. 4). This is an understandable consequence of the loss of information to separate
474 full and reduced models when there are fewer data. Nonetheless, it is an undesirable property,
475 just as the change in the unadjusted OLS R^2 is undesirable.

476 The poorer performance of all three R^2s for GLMM and PLOG relative to LMM and
477 PGLS in terms of partitioning sources of variation (Fig. 3) and precision (Fig. 4) is due to the
478 greater challenges of fitting discrete data. This will affect the R^2s differently if they are
479 differently sensitive to the fitting. R^2_{ls} is calculated from fitted variances in a model; R^2_{ce} is
480 calculated from the fitted values of Y_i; and R^2_{lr} is calculated from the likelihood. Therefore, the
481 three R^2s will be sensitive to the precision with which each of these attributes is estimated. For
482 GLMMs, the precision of R^2_{lr} was slightly greater than the other two R^2s, although this did not
483 appear to be the case for PLOG.

484 Although it is hard to argue in favor of one R^2 over the others on the basis of their
485 performance in the simulations, R^2_{ls} has the advantage of producing $R^2_{glmm(c)}$ (Nakagawa &
486 Schielzeth 2013) as a special case when applied to LMMs and GLMMs. Furthermore, because
487 R^2_{ls} (like the other R^2s) is defined as a partial R^2, it makes application to LMMs and GLMMs
488 more flexible and general, allowing subsets of fixed and random effects to be analyzed, and also
489 more-complex structures like random slopes (Johnson 2014). Nonetheless, a partial $R^2_{glmm(c)}$
490 comparable to R^2_{ls} can easily be defined as

491 $$R^2_{glmm.partial} = 1 - \frac{1 - R^2_{glmm(c).full}}{1 - R^2_{glmm(c).reduced}} \quad (9)$$

492 Although the relationship between R^2_{ls} and $R^2_{glmm(c)}$ might argue in favor of R^2_{ls} over R^2_{ce} and R^2_{lr},
493 when applied to data simulated from LMM and GLMM, R^2_{ls} values calculated from fitted LMM
494 and GLMM were different from the values calculated from PGLS and PLOG fitted to the same
495 data (Table 2). This highlights a weakness of R^2_{ls}: a decision has to be made about how to scale
496 the covariance matrix $\mathbf{\Sigma}$ depending on the fitted model, and the resulting values of R^2_{ls} will
497 depend on this decision.
498

499 *Recommendations*

500 An ideal R^2 would make it possible to compare among different models and among
501 different methods used to fit the same model (Kvalseth 1985 properties of a good R2 #4 and #5).
502 R^2_{ls} and R^2_{ce} can be used for any model and fitting method that estimates the covariance matrix
503 (R^2_{ls}) and/or fitted values (R^2_{ce}); for example, they could be used to compare LMMs fit with ML
504 vs. REML, or binary phylogenetic models fit with ML (e.g., phyloglm; Ho & Ane 2014) or
505 quasi-likelihood (e.g., binaryPGLMM; Ives & Garland 2014). Nonetheless, R^2_{ls} and R^2_{ce} have a
506 disadvantage in terms of generality. For R^2_{ls} a decision must be made about how to scale the
507 covariance matrix $\mathbf{V}(\theta)$ (equation (3)), and for R^2_{ce} a decision has to be made about how values
508 of Y_i are predicted. The conventions I used for LMMs and PGLS differed. In contrast, R^2_{lr} is
509 restricted to models that are fit with ML estimation; however, if ML is used for fitting, then
510 values of R^2_{lr} can be compared across different types of models. This applies to any type of data
511 and model fit with ML estimation.

512 An ideal R^2 should also be intuitive (Kvalseth 1985 property #1). However, intuitive is in
513 the eye of the beholder. R^2_{ls} is the most similar to the OLS R^2, which grounds R^2_{ls} in the familiar
514 and intuitive OLS framework. R^2_{ce} predicts the data from covariances estimated in the model,

and therefore could be viewed as the most intuitive way to relate the variance explained by regression coefficients (fixed effects) to that explained by variance parameters (random effects). R^2_{lr} is also related to the OLS R^2: in LMMs and PGLS, R^2_{lr} only differs from R^2_{ls} by the way in which the covariance matrix $\mathbf{V}(\theta)$ (equation (3)) is scaled, and this provides a link between R^2_{lr} and the OLS R^2 through R^2_{ls}. This said, however, I suspect that different researchers would rank the intuitiveness of R^2_{ls}, R^2_{ce}, and R^2_{lr} differently.

R^2s 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^2s be judged as a summary statistic? I think not. All the R^2s 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 dataset 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^2s are best treated as inferential statistics, that is, as functions of a data-generating process that are themselves random variables (Cameron & Windmeijer 1996; Nakagawa & Schielzeth 2013). As an inferential statistic, R^2_{lr} most directly ties to hypothesis testing between full and reduce models using a likelihood ratio test. For me, this tips the balance to favor R^2_{lr} over the others.

ACKNOWLEDGMENTS

I thank Ted Garland, Daijiang Li, Shinishi Nakagawa, Eric Pedersen, and Joe Phillips for wonderfully insightful comments that helped to clarify this article.

FUNDING

Financial support came from the US National Science Foundation, DEB-LTREB-1052160 and DEB-1240804.

REFERENCES

Barton, K. (2016) MuMIn: Multi-model inference.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014) lme4: Linear mixed-effects models using Eigen and S4.

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.

Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association,* **88,** 9-25.

Burnham, K.T. & Anderson, D.R. (2002) *Model selection and inference: a practical information-theoretic approach,* Second edn. Springer, New York, New York.

Buse, A. (1973) Goodness of fit in generalized least-squares estimation. *American Statistician,* **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. *Journal of Business & Economic Statistics,* **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. *Journal of Econometrics,* **77,** 329-342.

Cox, D.R. & Snell, E.J. (1989) *The analysis of binary data*. Chapman and Hall, London, UK.

Cragg, J.G. & Uhler, R.S. (1970) Demand for automobiles. *Canadian Journal of Economics,* **3,** 386-406.

Desdevises, Y., Legendre, P., Azouzi, L. & Morand, S. (2003) Quantifying phylogenetically structured environmental variation. *Evolution,* **57,** 2647-2652.

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. *Statistics in Medicine,* **27,** 6137-6157.

Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist,* **125,** 1-15.

Gelman, A. & Hill, J. (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York, NY.

Grafen, A. (1989) The phylogenetic regression. *Transactions of the Royal Society of London B, Biological Sciences,* **326,** 119-157.

Hadfield, J.D. (2015) Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods in Ecology and Evolution,* **6,** 706-714.

Ho, L.S.T. & Ane, C. (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology,* **63,** 397-408.

Housworth, E.A., Martins, E.P. & Lynch, M. (2004) The phylogenetic mixed model. *American Naturalist,* **163,** 84-96.

Ives, A.R. & Garland, T. (2010) Phylogenetic logistic regression for binary dependent variables. *Systematic Biology,* **59,** 9-26.

Ives, A.R. & Garland, T., Jr. (2014) Phylogenetic regression for binary dependent variables. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (ed. L.Z. Garamszegi), pp. 231-261. Springer-Verlag, Berlin Heidelberg.

Ives, A.R. & Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs,* **81,** 511-525.

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. *Journal of Applied Statistics,* **44,** 1086-1105.

Johnson, P.C.D. (2014) Extension of Nakagawa & Schielzeth's R-GLMM(2) to random slopes models. *Methods in Ecology and Evolution,* **5,** 944-946.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. & Lee, T.-C. (1985) *The theory and practice of econometrics,* Second edn. John Wiley and Sons, New York.

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*. New Prairie Press, Manhatan, Kansas.

Kvalseth, T.O. (1985) Cautionary note about R2. *American Statistician,* **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. *Physiological and Biochemical Zoology,* **81,** 526-550.

Liu, H.H., Zheng, Y. & Shen, J. (2008) Goodness-of-fit measures of R(2) for repeated measures mixed effect models. *Journal of Applied Statistics,* **35,** 1081-1092.

Maddala, G.S. (1983) *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge, UK.

Magee, L. (1990) R2 measures based on wald and likelihood ratio joint significance tests. *American Statistician,* **44,** 250-253.

Martins, E.P. & Hansen, T.F. (1997a) Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist,* **149,** 646–667. Erratum 153:448.

Martins, E.P. & Hansen, T.F. (1997b) Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist,* **149,** 646-667.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models,* 2 edn. Chapman and Hall, London.

Menard, S. (2000) Coefficients of determination for multiple logistic regression analysis. *American Statistician,* **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. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution,* **4,** 133-142.

Orelien, J.G. & Edwards, L.J. (2008) Fixed-effect variable selection in linear mixed models using R-2 statistics. *Computational Statistics & Data Analysis,* **52,** 1896-1907.

Pagel, M. (1997) Inferring evolutionary processes from phylogenies. *Zoologica Scripta,* **26,** 331-348.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics,* **20,** 289-290.
Paradis, E. & Paradis, E. (2012) *Analysis of Phylogenetics and Evolution with R Introduction*.
Petersen, K.B. & Pedersen, M.S. (2012) The matrix cookbook. Technical University of Denmark.
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. *Statistics in Medicine,* **22,** 3527-3541.

SUPPLEMENTARY MATERIAL

Appendix 1: Details about the implementation of the R^2s.

Appendix 2: More comparisons among the R^2s in the R package rr2.

supplementary files: source R scripts for computing R^2_{ls}, $R^2_{lr,}$ and R^2_{ce} with examples.

Table 1: Simulation example of sprint speed regressed on log body mass (x_1) and log hind limb length (x_2). For the simulation, the regression coefficients for log(*body size*) and log(*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 normal random variable (right table), in both cases with variance 1.

Phylogenetic signal in x_2

coefficient	estimate	P
intercept	-0.46	0.50
x_1	0.79	0.006
x_2	0.55	0.012
λ	1	

N = 30, logLik = -17.76

reduce model	R^2_{ls}	R^2_{ce}	R^2_{lr}
x_2=0	0.21	0.23	0.21
λ=0	0.69	0.89	0.72
x_1=x_2=λ=0	0.79	0.93	0.81

No phylogenetic signal in x_2

coefficient	estimate	P
intercept	-0.49	0.46
x_1	0.75	0.008
x_2	0.55	<<0.001
λ	1	

N = 30, logLik = -17.12

reduced model	R^2_{ls}	R^2_{ce}	R^2_{lr}
x_2=0	0.71	0.89	0.79
λ=0	0.71	0.90	0.74
x_1=x_2=λ=0	0.88	0.96	0.90

Table 2. Illustrative simulation comparing R^2s for data simulated from a LMM and fitted with LMM and PGLS, and for data simulated from a binary GLMM and fitted with GLMM and PLOG. Data were simulated from equation (7) with 10 levels of 10 observations each (n = 100), a random effect with variance 2, a normally distributed x with mean 0 and variance 1, and for the LMM a residual error variance of 0.5. PGLS and PLOG were fit by converting the covariance matrix given by the random effects into a phylogeny (Fig. 1). Columns labeled "Mixed" and "Phylo" correspond to LMM and GLMM, and PGLS and PLOG, respectively. For the LMM, the adjusted R^2_{adj} was computed from the OLS model fix by treating the random effect as a fixed effect with 10 levels.

		R^2_{lr}		R^2_{ce}		R^2_{ls}		R^2_{adj}
		Mixed	Phylo	Mixed	Phylo	Mixed	Phylo	
LMM								
	total	0.916	0.916	0.955	0.945	0.950	0.853	0.955
	partial x	0.722	0.722	0.764	0.763	0.764	0.490	0.764
	partial θ	0.913	0.913	0.954	0.943	0.949	0.848	0.954
GLMM								
	total	0.317	0.317	0.471	0.468	0.591	0.550	
	partial x	0.141	0.141	0.161	0.159	0.343	0.297	
	partial θ	0.298	0.298	0.461	0.459	0.577	0.535	

Figure legends

Fig. 1. Depictions of the covariance matrices from PGLS, LMM, PLOG, and GLMM. In the PGLS with a Pagel's λ branch-length transform, the covariance matrix is $\mathbf{\Sigma}(\lambda) = \lambda\mathbf{\Sigma}_{\mathrm{BM}} + (1-\lambda)\mathbf{I}$, in which λ determines the strength of phylogenetic signal in the residual errors. In the covariance matrix for LMMs, $\mathbf{\Sigma}(\sigma^2_b) = \sigma^2_b\,\mathbf{\Sigma}_b + \mathbf{I}$, the variance of the random effect σ^2_b is scaled against the variance of the residual errors. For PLOG, phylogenetic signal enters the model as a covariance matrix $\sigma^2_o\mathbf{\Sigma}_{\mathrm{BM}}$, but there is additional variance σ^2_w that depends on the difference between observed and predicted values of Y_i which varies for each data point. Similarly, for GLMMs the variance of the random effect is given by $\mathbf{\Sigma}(\sigma^2_b) = \sigma^2_b\,\mathbf{\Sigma}_b$, and there is additional variance σ^2_w owing to the discreteness of the data.

Fig. 2. Results for LMM, PGLS, GLMM, and PLOG simulations giving R^2_{ls}, R^2_{ce}, R^2_{lr}, and the OLS R^2_{adj} versus the log likelihood ratio (LLR) between full model and reduced model containing only an intercept. All simulated data had 100 samples. For LMM, the simulation model (equation (7)) 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; for PLOG $\beta_1 = 0$ or 1.5, and $\theta = 0$ or 2. The LMM was fit using lmer (Bates *et al.* 2014); the GLMM was fit using glmer (Bates *et al.* 2014); the PGLS was fit using phylolm (Ho & Ane 2014); and for PLOG LLR and R^2_{lr} were fit using phyloglm (Ho & Ane 2014), and R^2_{ls} and R^2_{ce} were fit using binaryPGLMM (Ives & Garland 2014). For reduced models without variance parameters, fitting was done using lm and glm.

680 **Fig. 3.** Results for LMM, PGLS, GLMM, and PLOG simulations giving partial values of R^2_{ls},

681 R^2_{ce}, R^2_{lr}, and R^2_{adj}. The partial R^2 for β_1 was calculated using the reduced model in which θ is

682 removed, and for the partial R^2 for θ the reduced model had β_1 removed. The simulated data and

683 fitting methods are the same as in figure 1.

684

685 **Fig. 4.** Results for LMM, PGLS, GLMM, and PLOG simulations showing means, 66% and 95%

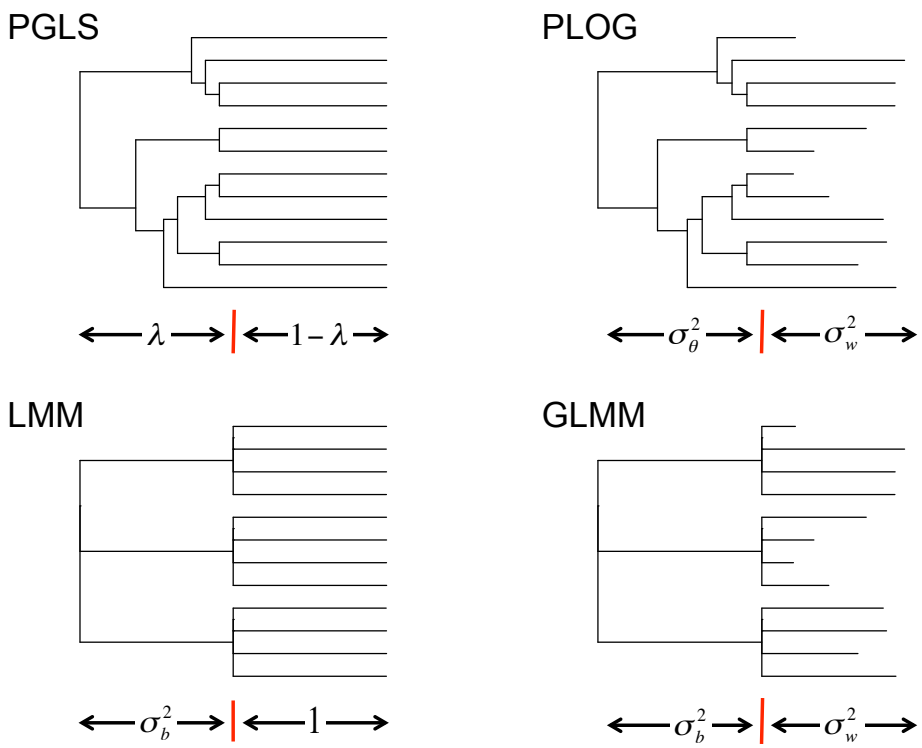
686 inclusion intervals for R^2_{ls}, R^2_{ce}, R^2_{lr}, and R^2_{adj} versus sample size. For all simulations 1000 data

687 sets were analyzed at each sample size. Parameter values were: LMM, $\beta_1 = 1$, $\theta = 1.5$; PGLS, β_1

688 $= 1$, $\theta = 0.5$; GLMM, $\beta_1 = 1.8$, $\theta = 1.8$; and PLOG, $\beta_1 = 1.5$, $\theta = 2$.

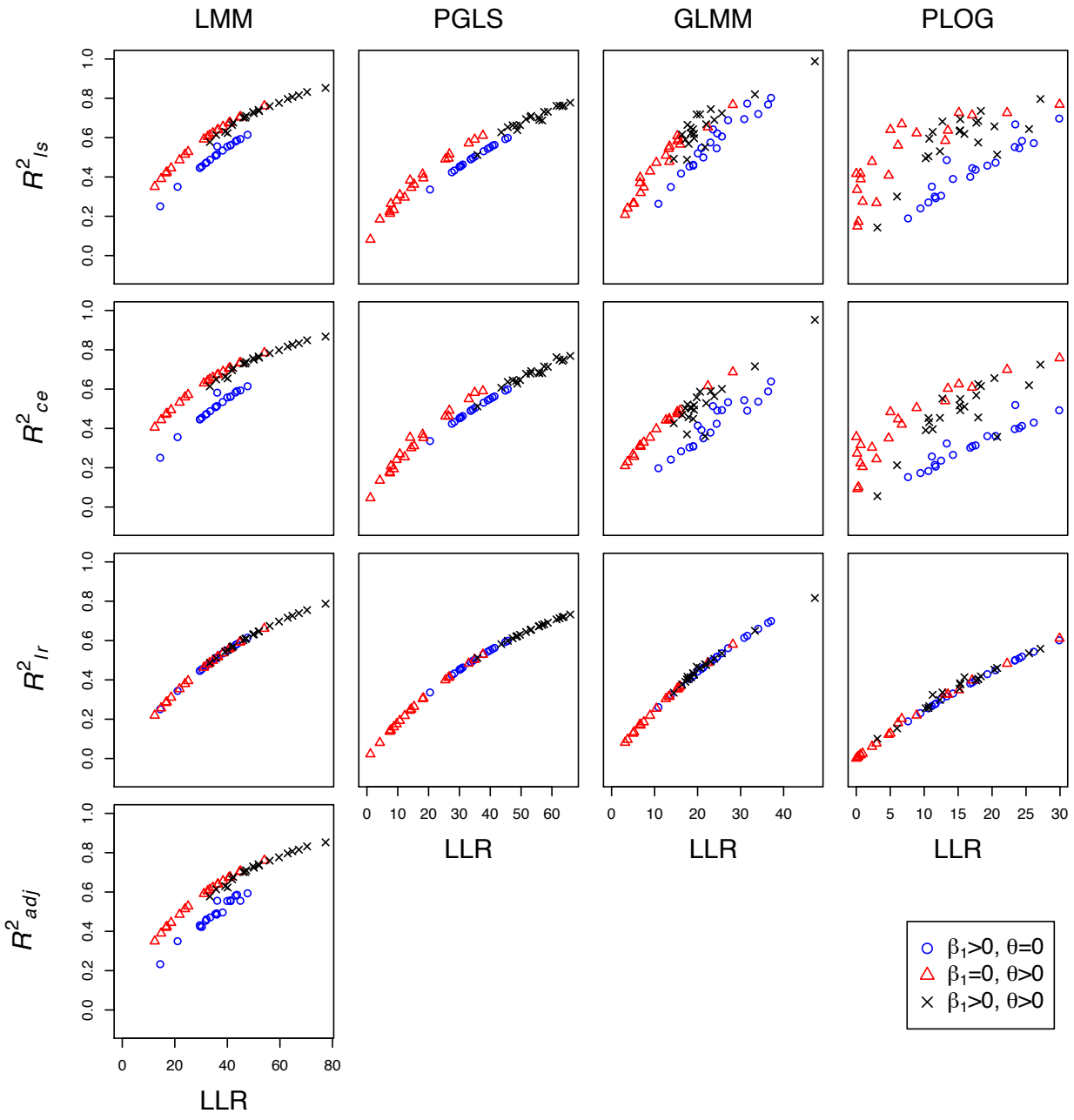
689

690 **Fig. 1**


PGLS
λ
$1-\lambda$
PLOG
σ_θ^2
σ_w^2
LMM
σ_b^2
1
GLMM
σ_b^2
σ_w^2

691

692 **Fig. 2**


LMM
PGLS
GLMM
PLOG
R^2_{ls}
R^2_{ce}
R^2_{lr}
R^2_{adj}
LLR
LLR
LLR
LLR
$\beta_1>0$, $\theta=0$
$\beta_1=0$, $\theta>0$
$\beta_1>0$, $\theta>0$

693

Fig. 3

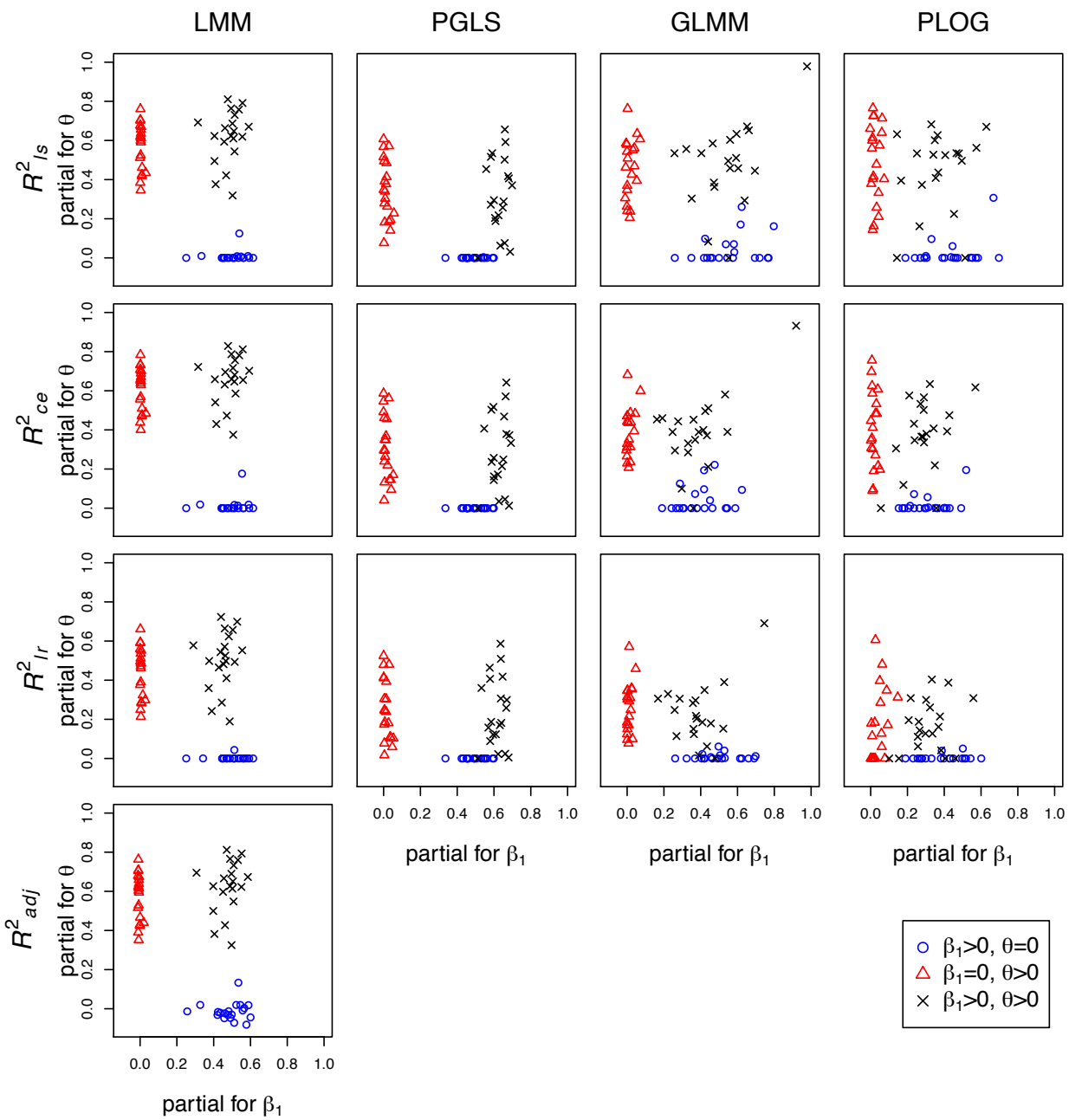
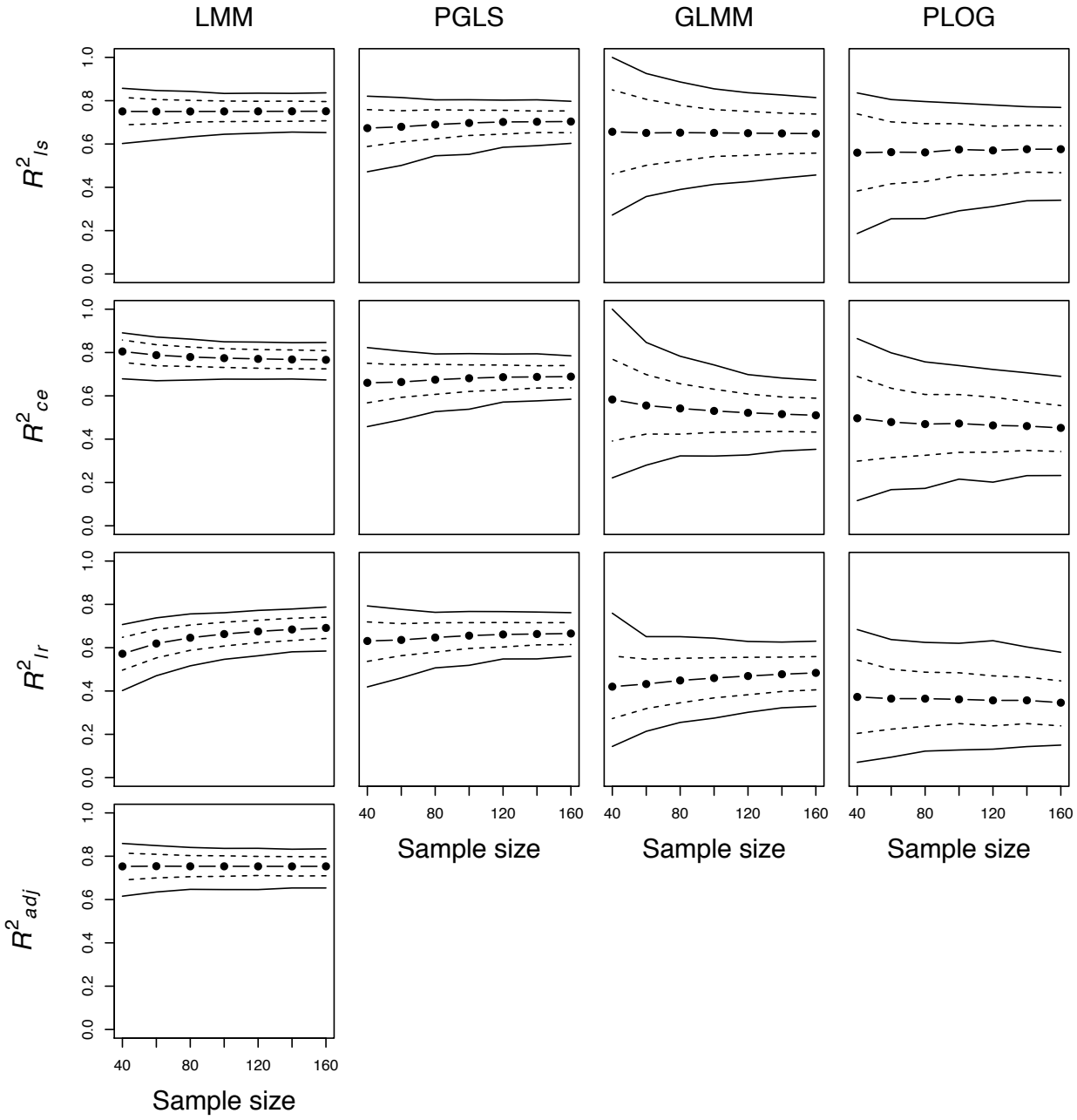

LMM
PGLS
GLMM
PLOG
R^2_{ls}
R^2_{ce}
R^2_{lr}
R^2_{adj}
partial for θ
partial for β_1
0.0
0.2
0.4
0.6
0.8
1.0
β_1>0, θ=0
β_1=0, θ>0
β_1>0, θ>0

Fig. 4


LMM
PGLS
GLMM
PLOG
R^2_{ls}
R^2_{ce}
R^2_{lr}
R^2_{adj}
0.0
0.2
0.4
0.6
0.8
1.0
40
80
120
160
Sample size