Chapter 9
Phylogenetic Regression for Binary Dependent Variables

Anthony R. Ives and Theodore Garland Jr.

Abstract We compare three methods for phylogenetic regression analyses designed for binary dependent variables (traits with two discrete states) both with each other and with "standard" methods that either ignore phylogenetic relationships or ignore the binary character of the dependent variable. In simulations designed to reveal statistical problems arising in different methods, PLogReg (Ives and Garland 2010) performed better than PGLMM (Ives and Helmus 2011) and MCMCglmm (Hadfield 2010) to identify phylogenetic signal in the absence of independent variables; PLogReg also outperformed a standard method for detecting phylogenetic signal in binary data, ancestral character estimation (Schluter et al. 1997; Pagel 1994). All three phylogenetic methods performed similarly for identifying relationships with a continuously valued independent variable x, with all methods having at most moderately inflated Type I error rates, and MCMCglmm having slightly greater power. In contrast, standard logistic regression that ignores phylogeny had seriously inflated Type I errors when x had phylogenetic signal. Perhaps surprisingly, phylogenetic regression that ignored the binary nature of the dependent variable, RegOU (Lavin et al. 2008), performed as well or better than the other methods, at least for larger sample sizes ($\geq$64 species), although this approach does not result in a model that can be used to simulate data (e.g., for bootstrapping). We also apply the methods to a data set describing whether antelopes fight or flee versus hide from predators as a function of their group size (Brashares et al. 2000). We end with rough guidelines for analyzing binary dependent variables, with the main recommendation being that multiple methods and simulations should be used to give confidence in the statistical results.

A. R. Ives (✉)
UW-Madison, Madison, WI 53706, USA
e-mail: arives@wise.edu

T. Garland Jr.
UC-Riverside, Riverside, CA, USA
e-mail: theodore.garland@ucr.edu

L. Z. Garamszegi (ed.), *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, DOI: 10.1007/978-3-662-43550-2_9,

9.1 Introduction

Generally speaking, comparative data from phylogenetically related species (or higher taxa) cannot be analyzed using standard statistical procedures because mean values for a set of species (or their residuals from a statistical model) are unlikely to be independent and identically distributed (Felsenstein 1985; Garland et al. 1992; Harvey and Pagel 1991). Computer simulation studies have shown, for example, that ignoring phylogenetic correlations often leads to Type I errors, rejecting a null hypothesis that is in fact true (Grafen 1989; Martins and Garland 1991; Diaz-Uriarte and Garland 1996; Martins et al. 2002). The resulting falsely low P-values in statistical tests may lead to seriously wrong conclusions in comparative studies. Thus, use of statistical methods that incorporate phylogenetic information is essential. For continuously valued traits, a growing number of methods and associated software have been developed which incorporate the possibility of phylogenetic correlations among taxa.

Traits with non-Gaussian distributions are surprisingly more difficult to analyze than continuously valued traits when they occur as dependent (response) variables. This difficulty arises in part because most non-Gaussian distributions have means and variances that are not separable. In other words, the variance of the distribution depends on the mean. This is different from the familiar Gaussian distribution in which one parameter gives the mean and a second gives the variance. For example, for a binomial distribution with probability p and number of trials n, the mean is np and the variance is $np(1 - p)$, so it is not possible to change n and/or p in any way that changes the mean without also changing the variance. This complication can be approximately addressed by transforming data. For example, if binomial data are divided by $\sqrt{n}$, then the variance is $p(1 - p)$ which can be changed independently of the mean $\sqrt{n}p$, thus allowing the data to be fit with a standard regression model that assumes a Gaussian distribution of residuals. This is still an approximation, however, because the regression model assumes that the data can take any value, not just integers between 0 and n. Therefore, the model does not really fit the process that underlies the data and hence cannot be used to simulate data with the same statistical properties. An inability to simulate data makes it difficult to determine the performance of statistical methods designed to analyze such data and generally precludes the use of methods based on simulations to obtain null distributions of test statistics. To solve these problems, the last 20 years have seen huge advances in methods designed to analyze non-Gaussian distributions, and these have spread into phylogenetic comparative methods.

One common type of inherently non-Gaussian response variable is binary, such as when an organism either does or does not possess a particular phenotypic trait (e.g., wings). Three methods have been developed for phylogenetically informed analysis of binary dependent variables: (i) phylogenetic logistic regression, implemented in a MATLAB program named PLogReg (Ives and Garland 2010), (ii) generalized linear mixed models with frequentist estimation, PGLMM (Ives and Helmus 2011), and (iii) generalized linear mixed models with Bayesian

estimation, MCMCglmm (Hadfield 2010). These methods address the dual challenges of formulating an appropriate statistical model and estimating parameters from the model. By accounting for the binary nature of the dependent variable, the hope is that these methods maximize statistical power, that is, the ability to identify parameters that differ statistically from zero or some other value of interest. Power is a major concern, because binary variables contain relatively little information as compared to continuously valued variables (Ives and Garland 2010); for binary variables, information is only available in the form of zeros and ones, with no finer gradation between values. Therefore, there is a premium on methods for binary data that have high statistical power.

When these three phylogenetic methods are used without independent variables, they become tests of phylogenetic signal (sensu Blomberg et al. 2003) in the response variable of interest (Chap. 5). Here, we first compare the three methods in their abilities to identify phylogenetic signal, including two additional "standard" methods: (iv) phylogenetic regression ignoring the binary nature of the dependent variable and assuming an Ornstein–Uhlenbeck model of residual trait variation (Chap. 15), RegOU (Lavin et al. 2008) and (v) maximum likelihood (ML) estimation of discrete traits evolving along a phylogenetic tree using ancestral character estimation, ACE (Pagel 1994; Schluter et al. 1997).

We also compare (i) PLogReg, (ii) PGLMM, and (iii) MCMCglmm for the simple regression case of a single continuously valued independent (predictor) variable, focusing on their abilities to estimate and perform statistical tests on a regression coefficient. In this comparison, we additionally include (iv) RegOU and (v) non-phylogenetic logistic regression with a Firth correction, Logistf from **logistf** {R} (Heinze et al. 2013). We include RegOU because it runs quickly in various software implementations and it might perform adequately in many cases, even though it ignores the binary nature of the dependent variable. We include Logistf because it is a standard method and illustrates the mistakes that can be made by not accounting for phylogeny.

Additional methods that address related statistical problems are not included in our comparisons. For example, Pagel (1994) presents a method for estimating the correlation between two phylogenetically related binary traits. We excluded this approach because here we focus on regression rather than correlation. Felsenstein (2012) presents an estimation procedure for threshold models that can test the correlation between both binary and continuously valued dependent variables. When applied to only binary variables, this method should be similar to MCMCglmm (Hadfield and Nakagawa 2010), so we do not include it here. Finally, note that the methods we analyze are for binary dependent variables. Binary independent variables with continuously valued dependent variables present no special problems for the existing phylogenetic methods for regression and ANOVA or ANCOVA (Garland et al. 1993; Lavin et al. 2008; Dlugosz et al. 2013; Revell 2012; Rezende and Diniz 2012).

A subtext to this chapter is that there is no single "correct" way to perform phylogenetic regressions with binary dependent variables. All of the methods have strengths and weaknesses that must be balanced for a specific data set and

question. The lack of a single best method is not uncommon in statistics when confronted with data arising from complicated processes. The best that can be done is to use different methods, know which is likely to perform best under circumstances resembling those of the data under analysis, and use this knowledge to help select the "best" result. We hope this chapter provides a rough guide for doing this. Despite providing a guide, we hope that we also show that there is no substitute for careful analysis of any complicated set of data, including applying multiple methods and simulations to rigorously confirm the results.

Below, we first give descriptions of the methods. We then perform simulations to compare them, first for the case without independent variables where the methods become tests for phylogenetic signal and then for a simple regression with one independent variable. These simulations are by no means exhaustive; we use them more to illustrate the issues that arise in statistical analyses of binary data and possible ways to address these issues, rather than to give recommendations for which method to use for a particular data set and question. Indeed, one of our main points is that there is no best method for all situations, so you should use our simulations as guides to the types of simulations you should perform. Finally, we apply the methods to a comparative data set which was analyzed previously using phylogenetic logistic regression (Ives and Garland 2010). The online practical material (http://www.mpcm-evolution.org) presents this analysis as a tutorial and also the new code in R that performs the PGLMMs.

9.2 Description of Statistical Models and Estimation Procedures

In building a statistical model, it is often valuable to consider the underlying processes that might generate a data set. For example, for continuously valued traits (e.g., body mass or length), a simple Brownian motion (BM) model (a random walk in continuous time, Chap. 5) could be used to model phenotypic changes that occur in a population experiencing no selection but nonetheless evolving because of random mutation and genetic drift (Felsenstein 1985; Freckleton et al. 2002; Blomberg et al. 2003). Equally, however, the BM model could describe species under strong selection that track the environment instantaneously provided the environment itself changes randomly according to a BM process. The logical conclusion from this recognition is that the observed pattern of trait values among extant species does not necessarily give a lot of information about the processes generating this pattern (Revell et al. 2008). Nonetheless, building a statistical model under a specific evolutionary assumption can lead to a model with useful statistical properties. For example, the BM model can be modified to accommodate stabilizing selection using the Ornstein–Uhlenbeck process, borrowed from physics (Felsenstein 1988; Garland et al. 1993; Martins and Hansen 1997; Chap. 15). In this case, varying the strength of stabilizing

selection varies the strength of phylogenetic correlations between species, thereby producing a statistical model that can be used to estimate phylogenetic signal (Martins and Hansen 1997; Hansen 1997; Blomberg et al. 2003).

For binary traits, no single generally applicable evolutionary or statistical model exists. The lack of a uniquely suitable model is common in statistics, at least after graduating beyond the simplest problems. It is not an issue restricted to the realm of phylogenetically informed statistical procedures. For example, for regression with a binary response variable in the absence of phylogenetic considerations, both logistic and probit models are used routinely, each with its own advantages and disadvantages (Gelman and Hill 2007). We now describe the five approaches that we will compare. We give largely heuristic descriptions of the methods, as opposed to technical descriptions, because these methods are already in the literature.

9.2.1 Phylogenetic Logistic Regression (PLogReg)

The overall structure of PLogReg (Ives and Garland 2010) looks like that of standard logistic regression. A single trait Y can take only one of two values (0 or 1) with probability p, which itself depends on the independent (predictor) variable x. Multiple independent variables can be used, and they can be binary, multistate (coded into a set of 0–1 dummy variables), or continuously valued. A formal specification of the model with a single continuously valued independent variable x is

$$\begin{aligned} \Pr(Y = 1) &= p \\ \text{logit}(p) &= b_0 + b_1 x \\ \text{cov}(Y) &= \mathbf{V}(p, a) \end{aligned} \tag{9.1}$$

Here, b_0 and b_1 are regression coefficients, and logit(p) is the logit function $\log(p/(1-p))$ that maps any value of p in the interval (0, 1) onto values of $b_0 + b_1 x$ between $-\infty$ and ∞. Thus, for a given value x, the probability that $Y = 1$ is $p = \text{logit}^{-1}(b_0 + b_1 x)$. No term is included for the "residual variation" because for a binary stochastic process, the variance is determined by the mean; specifically, the variance equals $p(1-p)$. Thus, this is different from a conventional least-squares regression equation, in which an explicit vector of residual deviations from the predicted values is an inherent part of the statistical model. Even though the variance of the binary dependent variable is specified by the mean, the anticipated covariances can be positive as specified by the covariance matrix $\mathbf{V}(p, a)$. These covariances represent phylogenetic signal, that is, the lack of independence among data points Y caused by taxa having experienced a shared evolutionary history prior to the speciation event(s) that begat them. Because the variances depend on the mean p, so too do the covariances, and this presents complications in the statistical model building and estimation not only for the

regression coefficients b_0 and b_1 but also for the parameter that gives the strength of phylogenetic signal, which in PLogReg is called a (Ives and Garland 2010). Although Eq. (9.1) is written with a single independent variable x, PLogReg can accommodate multiple independent variables, and when used without any independent variables (with only regression parameter b_0), PLogReg becomes a method to estimate phylogenetic signal given by the parameter a.

The evolutionary model used to build PLogReg assumes that Y evolves along a phylogenetic tree. There is a constant "instantaneous" probability of the trait switching from 0 to 1 or from 1 to 0, so the more time that elapses, the greater the chance of a switch occurring. The branch lengths of the phylogenetic tree scale the time between nodes, so the probability of a switch occurring between nodes increases with the branch length between nodes. Repeated switches can occur, and an even number of switches results in no difference in Y between two nodes (branching points) on the tree.

Phylogenetic signal arises because any two related species will have the same trait value at their nearest shared ancestral node, just before the speciation event. Depending on the switching rate for the trait in question, these two daughter species will be more or less likely to retain that ancestral state. If the switching rate is very low, then both daughters will likely retain the ancestral state, resemble each other, and hence provide evidence of phylogenetic signal. The overall switching rate given by the parameter a measures the strength of phylogenetic signal in trait Y. The parameter a is scaled so that larger values of a correspond to greater phylogenetic signal (lower switching rates). Although mathematically a can take any real value, numerically PLogReg limits values of a to range between -4 (no signal) and 4 (very strong signal).

Even though greater phylogenetic signal occurs for larger a, it may become very difficult to test for this signal statistically (Ives and Garland 2010). For example, in the extreme case of very strong phylogenetic signal (low switching rates), all taxa will likely share the same trait value (0 or 1), so there is no variation with which to test for statistical significance. This leads to the somewhat counterintuitive expectation that the power to detect phylogenetic signal occurs at intermediate strengths of the signal. Again, this is different from continuously valued traits where, despite limiting the divergence between taxa, strong phylogenetic signal can nonetheless be (strongly) detected in what variation in trait values is observed (Revell et al. 2008).

The process model for trait evolution just described does not involve any independent variables. PLogReg introduces an independent variable x after the evolutionary process establishes phylogenetic correlations in the values of Y among taxa. To model the effects of independent variables, starting with the values of Y following evolution up the phylogenetic tree, the model assumes that these values then rapidly evolve toward 0 or 1 depending on the value of x, independently for each species. In other words, the part of evolution of Y which is driven by the value of x does not depend on phylogeny. Although this is not a realistic model for many scenarios describing the evolution of dependent variables,

it makes sense statistically, because it ensures that if there is no phylogenetic signal in Y, then the model degenerates to standard logistic regression; in general, it is desirable to have phylogenetic methods give their conventional counterparts when there is no phylogenetic signal (Blomberg et al. 2003). This is an example of trade-offs that must sometimes be made for statistical necessity: Although it might not be evolutionarily plausible, the model underlying PLogReg leads to useful statistical properties, and the model parameters can be statistically fit.

Although this model can be statistically fit, doing so is not easy for two reasons that make it impossible to use standard statistical approaches and readily available software. First, the likelihood function of the model is complicated. The likelihood function is central to parameter estimation; it gives the likelihood of observing the data, given values of the model parameters. Therefore, the maximum likelihood (ML) parameter estimates are computed as those that give the greatest likelihood. For parameter estimation in PLogReg, we used the statistical approach of quasi-likelihood functions, which for technical reasons are well suited for logistic regression (see Ives and Garland 2010). The second statistical issue is that ML estimation for standard (non-phylogenetic) logistic regression is biased; standard ML estimates are on average further from zero than they should be (Heinze and Schemper 2002). In the non-phylogenetic case, this bias can be largely corrected by penalizing the likelihood function as suggested by Firth (1993). We used a similar approach in PLogReg. Simulations showed that this improves the statistical properties of the estimates of PLogReg parameters (Ives and Garland 2010).

Simple diagnostics for determining the adequacy of models for binary dependent variables do not exist, especially for the phylogenetic case. Of course, it is always important to plot the data and fitted model, which can be instructive for identifying gross violations of model assumptions. Nonetheless, the best approach to assess the quality of parameter estimates is to perform a "parametric bootstrap" (Efron and Tibshirani 1993). A parametric bootstrap takes a fitted model and uses it to simulate (a large number of) data sets. The parameters are then reestimated for each of the simulated data sets. Some deep statistical theory shows that the distribution of the parameter values estimated from the simulated data sets approximates (i.e., approaches asymptotically) the theoretical distribution of the parameter estimates (Efron and Tibshirani 1993). This distribution can then be used to obtain confidence intervals and perform statistical tests regarding the parameters. For PLogReg, it is possible to approximate the distributions and confidence intervals of estimates of the regression parameters b_0 and b_1, but bootstrapping is the only effective way to obtain this information for the phylogenetic signal parameter a (Ives and Garland 2010).

It is also possible to perform a bootstrap to test null hypotheses, for example, that a regression coefficient is zero, H_0:$b_1 = 0$. This is done by fitting the data assuming $b_1 = 0$, simulating the model to produce bootstrap data sets, estimating b_1 for each simulated data set, and counting the number of values of b_1 for the simulated data sets that exceed the value of b_1 calculated from the data. In practice, however, this approach often gives very similar results to that of bootstrapping

using the observed value b_1 and testing H_0:$b_1 = 0$ using the bootstrapped confidence intervals of b_1.

An additional advantage of parametric bootstrapping is that it allows identification of bias in parameter estimates. If, for example, the mean of the parameter estimates from the simulated data sets is lower than the value obtained from fitting to the real data and consequently used to perform the simulations, then this would indicate downward bias in the estimates, including the estimate from the original data. It is possible to use this information to correct for bias; a value is picked that, when used in the simulations, produces a mean parameter value that matches the value computed from the data. However, we do not pursue this form of bootstrap bias correction here, instead using bootstrapping simply to identify the existence of bias. Although bootstrapping is useful, it is not a panacea, and there is no substitute for looking at the data and fitted model, and using different methods.

In the simulations, we used the MATLAB code for PLogReg (Ives and Garland 2010), although there is a fast version available in **phylolm** {R} (Ho and Ane 2014).

9.2.2 *Generalized Mixed Model with Frequentist Estimation (PGLMM)*

PGLMM (Ives and Garland 2010) is a "phylogenetic" implementation of a generalized linear mixed model (Gelman and Hill 2007; McCulloch et al. 2008; Bolker et al. 2009) for binary data. The PGLMM for a single independent variable x is

$$\begin{aligned} &\Pr(Y = 1) = p \\ &\text{logit}(p) = b_0 + b_1 x + \varepsilon \\ &\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{C}) \end{aligned} \tag{9.2}$$

Unlike PLogReg, PGLMM treats the probabilities p as random variables, with the distribution of logit(p) being given by a standard regression model that includes the random variable ε that contains phylogenetic information. The value of ε can be thought of as a continuously valued, phylogenetically inherited, but unmeasured trait. We assume that it evolves up the phylogenetic tree according to a BM evolutionary process (Chap. 5). This makes ε a Gaussian random variable with covariance matrix $\sigma^2\mathbf{C}$ in which diagonal elements c_{ii} are proportional to the branch lengths from the basal node of the phylogenetic tree to taxa i, and off-diagonal elements c_{ij} are proportional to the shared branch lengths between taxa i and j. The diagonal elements of $\mathbf{C}$ can be equal (for an ultrametric tree with contemporaneous tips) or unequal (e.g., for a tree with time-calibrated branch lengths in which some species became extinct in the distant past).

An evolutionary interpretation of PGLMM is that an underlying, unobserved continuously valued trait evolves up a phylogenetic tree. Then, for taxa at the tips of the tree, the value of this trait determines the probability of Y taking values 0 or 1. Therefore, two stochastic processes are in play: the evolution of the underlying continuous trait that gives the value of p and the choice of the value of Y given probability p. For example, a herbivorous insect might evolve increased production of a detoxifying enzyme (a continuous trait) that increases the chances that it can shift to use a new host plant containing high levels of the toxin (a binary trait: adoption or not of the host plant). Alternatively, the expression level of some gene might determine whether wings develop during ontogeny. The PGLMM model differs slightly from "threshold" models (Felsenstein 1988, 2012) in which the value of Y is determined strictly according to whether the underlying continuous trait exceeds a threshold value, although threshold and PGLMM models are broadly equivalent from a functional perspective (Hadfield and Nakagawa 2010).

In PGLMM, σ^2 measures phylogenetic signal. If $\sigma^2 = 0$, Eq. (9.2) reduces to standard logistic regression (McCullagh and Nelder 1989). If $\sigma^2 > 0$, then the phylogenetic covariances between values of ε lead to covariances between the values of Y. Unlike PLogReg, there are no covariances between values of Y other than those contained in $\mathbf{C}$ and generated by ε. A important statistical property of this model is that the variances (diagonal elements of $\mathbf{C}$) are redundant. For example, assume that the phylogenetic tree is a "star" with unconnected branches of equal length leading from basal node to each tip (terminal) taxon. The corresponding phylogenetic covariance matrix has identical diagonal elements and zero off-diagonal elements. Therefore, even if $\sigma^2 > 0$, no phylogenetic signal exists. As a consequence, the variances have no real effect in the model: The value of σ^2 does not affect the variance in the values of Y, because we know that the variance equals $p(1 - p)$. The only effect σ^2 has is to complicate the interpretation of the regression coefficients b_0 and b_1. Because the logit function is nonlinear, the expected value of p is not equal to $\text{logit}^{-1}(b_0 + b_1 x)$ when there is variation in ε. Therefore, the variance in ε determined by σ^2 will affect the estimates of b_0 and b_1.

Technically, estimating σ^2 when no phylogenetic signal exists represents a problem of statistical "identifiability" (Judge et al. 1985) because σ^2 and the regression coefficients cannot be estimated simultaneously: Different combinations of parameter values can give an identical fit of the statistical model to the data. Identifiability is a very general phenomenon that plagues statistical analyses when the model is improperly specified. A familiar example is that of collinearity of independent variables in standard linear regression: If two independent variables are perfectly correlated, then it is impossible to estimate regression coefficients for both. This problem technically goes away if the independent variables are even slightly less than perfectly correlated, although practically the problem remains for even moderately highly correlated independent variables (e.g., >0.7) unless sample sizes are large. Our particular identifiability problem is similar in that if any phylogenetic covariances (off-diagonal elements) are present in matrix $\mathbf{C}$, then it is technically possible to estimate σ^2 and the regression coefficients. Nonetheless, in practice, when phylogenetic signal is weak (off-diagonal elements of

matrix **C** are small), it will be hard to estimate σ^2 unless sample sizes (number of taxa) are very large. We will discuss statistical problems associated with identifiability when describing our simulation results.

An additional problem arises when comparing values of regression coefficients b_0 and b_1 in Eq. (9.2) for data sets that differ in phylogenetic signal and hence σ^2. Because σ^2 affects the estimates of b_1, apparent differences in regression coefficients among models could be caused by differences in the estimated strength of the relationship between Y and x or by differences in the magnitude of phylogenetic signal. A simple way to at least approximately correct for this is provided by Hadfield (2012), following Diggle et al. (2004): If b_1 is estimated when $\sigma^2 > 0$, then the value that would have been estimated in the absence of variance in ε is approximately $b_1(1 + c^2\sigma^2)^{-0.5}$ where $c = (16/15)\,(3^{1/3}/\pi)$. Thus, if all estimates are corrected by a factor $(1 + c^2\sigma^2)^{-0.5}$, then the values of the regression coefficients can be compared more directly. Note that although this correction facilitates comparisons among estimates for the PGLMM model (Eq. 9.2), the structure of this model is different from that of the PLogReg model, and therefore, the regression parameters are not expected to have exactly the same values for a given data set, even though both models provide valid estimates of the same relationship between Y and x.

In principle, various approaches can be used for parameter estimation for this PGLMM, although none is easy to implement. For example, the software package ASReml can be configured for PGLMM (Jarrod Hadfield, pers. comm.). Here, we use the approach presented in Ives and Helmus (2011) for solving a more general formulation of PGLMM models designed for data sets containing the presence/absence of species from ecological communities as dependent variables. Parameter estimation involves combining penalized quasi-likelihood (PQL) and restricted maximum likelihood (REML) in a two-step process. This approach gives approximate standard errors for the regression coefficients from which confidence intervals can be calculated and statistical tests can be performed. Bootstrapping can and should also be performed as described above for PLogReg. We implemented PGLMM in MATLAB and provide a version in R in the online practical material (http://www.mpcm-evolution.org); a more general but harder-to-use function is also available in **picante** {R} (Kembel et al. 2010).

9.2.3 *Generalized Mixed Model with Bayesian Estimation (MCMCglmm)*

A very similar model to PGLMM can be implemented using MCMCglmm {R} (Hadfield 2010) with a Bayesian framework (Chap. 10). The model is

$$\begin{aligned}
&\Pr(Y = 1) = p \\
&\text{probit}(p) = b_0 + b_1 x + s + u \\
&\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{s}}^2 \mathbf{C}) \\
&\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{u}}^2 \mathbf{I})
\end{aligned} \tag{9.3}$$

This is identical to Eq. (9.2), except a probit rather than logit transform is used, and there is an additional random term u. In the general formulation of GLMMs that MCMCglmm is designed to analyze, the random variable u represents residual variation, and s is a "random effect" that captures hypothesized covariances in the data. This model is identical in structure to the GLMMs that can be analyzed using such popular programs as lmer in **lme4** {R}, although most programs (including lmer) for technical reasons restrict the structure of covariance matrix $\mathbf{C}$ to contain blocks of covariance terms rather than covariances specific (and potentially unique) to pairs of species; this means that a phylogenetic covariance matrix of arbitrary form cannot be used. MCMCglmm has no such restriction and is an easy-to-use program for investigating phylogenetic GLMMs.

MCMCglmm is a Bayesian statistical approach, and therefore, statistical results have a different interpretation than those produced by other methods we investigated (Chap. 10). Although long discussions in the literature argue the benefits of Bayesian versus frequentist approaches, most statisticians we know are comfortable using both, adopting Bayesian analyses mainly when the likelihood function does not lend itself to be handled easily using frequentist approaches. The growth in the use of Bayesian statistics has been fueled largely by the growth of inexpensive computing power, which allows the application of the Markov chain Monte Carlo (MCMC) algorithm to complex likelihood functions (Gelman et al. 1995).

MCMC Bayesian statistics, including MCMCglmm, approximate the distribution of a parameter estimate conditional on the observed data and an initial "prior" specification of the parameter distribution before information about the data is used. The mode of the distribution of parameter values from the Markov chain is then the "best" estimate of the parameter, and the spread of the distribution determines the credibility of the parameter estimate. Rather than confidence intervals, MCMCglmm generates credible intervals that give, for example, the range of values that a parameter takes with 95 % probability. Although Bayesian approaches do not give tests of hypotheses in the frequentist sense, the credible intervals give similar information. For example, if the 95 % credible interval of b_1 lies above and does not include zero, then we could say that the value of b_1 is greater than zero with 95 % credibility.

Owing to the structure of MCMCglmm, it is necessary to include the random variable u; we fixed the variance $\sigma_u^2 = 1$, while estimating the variance σ_s^2 as a measure of phylogenetic signal. The inclusion of u makes the model different from the PGLMM model of Eq. (9.2). Nonetheless, this is not a serious difference, because of the identifiability issue discussed for PGLMM (Sect. 9.2.2) when the off-diagonal elements in covariance matrix $\mathbf{C}$ are all zero; the variance in u will

affect the specific values of the estimates of the regression coefficients, but will have little effect on the fit of the model. To make the regression coefficients more comparable between models and data sets, we also correct the regression coefficients of the MCMCglmm model by a factor $(1 + c^2\sigma_s^2 + c^2\sigma_u^2)^{-0.5}$ to "take away" the effect of variances on the estimates of the regression coefficients (Hadfield 2012); this should make values more comparable to those from PGLMM.

To run MCMCglmm, it is necessary to make decisions about the structure of the model, the prior parameter distributions, and the run characteristics. In our simulations, we found that using a probit transform (Eq. 9.3) performed better than a logit transforms as used in PGLMM (Eq. 9.2). We followed the recommendations of Hadfield (2012, Sect. 8.0.8) to use slice sampling and expanded priors for variance parameters σ_s^2 and σ_u^2, choosing X^2 priors as suggested by Villemereuil et al. (2013). The MCMCglmm defaults were used for the length of the burn-in (3,000) and sampled (10,000) chain lengths, with 1/10 samples taken for the posterior parameter distributions. Diagnostics (Hadfield 2012) showed adequate chain mixing for these settings. Discussion of these assumptions is provided in the online practical material (http://www.mpcm-evolution.org).

9.2.4 *Regression with a Model-Dictated Branch-Length Transformation (RegOU)*

For continuously valued dependent variables, the current state of the art in phylogenetic analyses is to perform regression while simultaneously estimating the strength of phylogenetic signal using a branch-length transform (Chap. 15). This is possible in a variety of packages; for example, **ape** {R} (Paradis et al. 2004) provides phylogenetic correlation structures corresponding to different models of evolution that can be used in regression through the gls function of **nlme** {R}. As noted elsewhere, so-called PGLS methods are equivalent to phylogenetically independent contrasts in their simplest form and do not inherently include estimation of branch-length transforms (Lavin et al. 2008). Here, we use the module RegOU in the Regressionv2.m (Lavin et al. 2008) MATLAB program to perform equivalent analyses under the assumption that evolution follows an Ornstein–Uhlenbeck process, intended to mimic stabilizing selection (Felsenstein 1988; Garland et al. 1993; Chap. 15). The strength of stabilizing selection given by the parameter d determines the phylogenetic signal estimated to exist in residuals. In the absence of stabilizing selection ($d = 1$), RegOU gives BM evolution, while stronger stabilizing selection (smaller d) erases evolutionary memory and reduces phylogenetic signal. When $d = 0$, phylogenetic signal disappears, and the estimated model parameters are then identical to those obtained by an ordinary least-squares analysis, which in effect assumes a star phylogeny with no hierarchical structure. An advantage of integrating a branch-length transform into the

regression analysis is that an a priori decision about the existence of phylogenetic signal is unnecessary; instead, it is estimated along with the regression coefficients—the data are allowed to tell their own story (Lavin et al. 2008).

We used RegOU even though we know the binary data Y violate the RegOU assumption that the dependent variable takes continuous values. Although one could object to applying a statistical model that is known to be wrong, all statistical models are likely to be wrong in some way to some degree, and the assumption of RegOU that Y be continuous might not make RegOU useless. By the central limit theorem, even a process as discrete as flipping a coin will converge to a Gaussian distribution if repeated enough times. With a sufficient number of taxa, RegOU might be adequate to identify, for example, the existence of an effect of independent variable x on Y. Because RegOU does not produce a model with discrete outcomes, however, the values of the regression coefficients are difficult to interpret. Furthermore, it is not possible to use a fitted model to simulate data, and therefore, it is not possible to perform bootstrapping. Nevertheless, RegOU might perform reasonably well for hypothesis testing (i.e., yield reasonable Type I errors and power).

9.2.5 Ancestral Character Estimation (ACE)

Pagel (1994) presented a rapid method for computing the exact likelihood for a model of two binary traits evolving along a phylogenetic tree in a correlated fashion. Using the same approach (Schluter et al. 1997) but applied to a single trait gives a method for estimating phylogenetic signal for a single binary trait. For a single trait, the assumed evolutionary process is essentially identical to the first part of the PLogReg model that generates phylogenetic correlations (Sect. 9.2.1). The trait Y has an instantaneous probability of changing from 0 to 1 and from 1 to 0, and the longer the branch lengths between nodes, the greater the probability of switches (and back-switches). The program ACE in **ape** {R} (Paradis et al. 2004) computes the ML of this process while fitting the transition rate parameters.

Because ACE does not lead to a model that can be used for regression with an independent variable x, we only use it to detect phylogenetic signal. As with PLogReg, higher transition rates reduce phylogenetic signal and therefore could be used to assess phylogenetic signal. Nonetheless, this does not lead to a clear test of whether phylogenetic signal is present, because the threshold for the transition rates above which we should declare no signal is unclear. Therefore, we used the following procedure. We used Grafen's (1989) rho branch-length transform (that is easily implemented using the compute.brlen function in **ape** {R}) to modify a given topology and hence produce trees representing different degrees of expected phylogenetic signal. As rho approaches zero, the tree becomes a star, with no phylogenetic covariances between taxa; in compute.brlen, rho can never equal zero, so we used a minimum values of rho $= 10^{-5}$ instead of zero. (Note that this restriction on the lower limit of rho is particular to compute.brlen; other programs

allow rho to become zero [e.g., DOS PDTREE and the PDAP module of Mesquite].) Larger values of rho correspond to greater shared branch lengths and hence greater phylogenetic structure. We used ACE to calculate the ML of the model over the possible transition rates for a given value of rho and then selected that value of rho giving the greatest likelihood. Thus, a test for the existence of phylogenetic signal is whether the rho giving the greatest likelihood is statistically greater than zero. ACE can analyze several models of evolution, such as allowing a different rate for the 0 to 1 transition than for the 1 to 0 transition. However, in preliminary simulation experiments, we found that the simplest model assuming transitions are symmetrical gave the greatest power to detect phylogenetic signal.

9.2.6 *Standard Logistic Regression with a Firth Correction (Logistf)*

To compare with a non-phylogenetic analysis, we used standard logistic regression with a Firth correction to reduce bias in the estimates (Ives and Garland 2010) as implemented by **logistf** {R} (Heinze et al. 2013). We only used this model in simulations including an independent variable x.

9.3 Method Comparison Using a Simulation Model

We performed simulations both without and with an independent variable x. In the absence of x, the phylogenetic models become tests of phylogenetic signal. To generate simulation data for comparisons among models, we chose a model similar to that underlying PLogReg, although differing slightly in the manner of generating phylogenetic signal. We first simulated switches in the value of Y along a phylogenetic tree at a rate given by $a = 0$, which gives phylogenetic signal comparable in magnitude to BM evolution of continuous traits (Ives and Garland 2010). However, rather than vary the value of a to change the strength of phylogenetic signal, instead, we performed an OU transform on the phylogenetic tree (following the parameterization of Lavin et al. 2008), with $d = 0$ giving no phylogenetic signal (a star phylogeny) and increasing values of d giving greater phylogenetic signal (see Sect. 9.2.4 above and Chap. 15 for a description of the OU branch-length transform). For the case with an independent variable x, again as in PLogReg, we assumed that a second phase of evolution occurred in which values of Y (at the tips of the tree) tended to switch toward 0 or 1 depending on the value of x and the regression coefficient b_1, and independently from the phylogeny.

To simulate values of x, we assumed values were chosen either with or without phylogenetic signal. To model phylogenetic signal in x, we assumed BM evolution

on the phylogenetic tree leading to a multivariate Gaussian distribution with mean 0, variance 1, and covariances proportional to the shared branch length for each tip species. To model x in the absence of phylogenetic signal, we chose values of x independently from a Gaussian distribution with mean 0 and variance 1. We considered only "symmetrical" ("balanced") phylogenetic trees; preliminary simulations showed little difference in results between symmetrical and highly asymmetrical "ladder" trees (see also Blomberg et al. 2003). Finally, when simulated data sets had $\geq$7/8 of the values of Y all zeros or all ones, we discarded them because they will contain little information, and a prudent researcher would not analyze them in the first place (cf. Diaz-Uriarte and Garland 1996, p. 45).

9.3.1 Phylogenetic Signal (Regression Without Independent Variables)

We first investigated the ability of five methods—PLogReg, PGLMM, MCMCglmm, RegOU, and ACE—to detect phylogenetic signal. We simulated 1,000 data sets for 64 species on a symmetrical phylogenetic tree subjected to an OU branch-length transform. We used d values in increments of 0.1 in the range of 0 to 0.8; although d can exceed unity, values >0.8 were deemed unnecessary, given the results obtained. For each of the five methods, we plotted the mean estimates of the respective phylogenetic signal parameters and their 66 and 90 % inclusion intervals (Fig. 9.1). The 66 % inclusion interval is comparable to ± 1 standard deviation. The 90 % inclusion interval gives an indication of the power of each method to reject the null hypothesis of no phylogenetic signal; because this hypothesis is one-sided, the lower boundary of the 90 % inclusion interval corresponds to 5 % of the simulated data having estimates of no phylogenetic signal. It is important to note that these inclusion intervals are not confidence intervals, because the simulations were produced under a model that differed from all of the statistical models. Nonetheless, the inclusion intervals give an indication of the probability that the point estimate of phylogenetic signal from each model is greater than zero. We did not perform simulations that involved estimating confidence intervals due to the computational intensity required; for example, if confidence intervals were computed for 1,000 bootstrap data sets for each of the 1,000 simulations, 1,000,000 estimations would be needed at each level of phylogenetic signal.

PLogReg was the best method for detecting signal, followed by PGLMM. Given the similarity between the PGLMM and MCMCglmm models (Eqs. 9.2 and 9.3), we were surprised that MCMCglmm did not perform better; even for the simulation $d = 0.8$, more than 12 % of the simulations resulted in MCMCglmm estimates of $\sigma_s^2 = 0$ indicating no phylogenetic signal (Fig. 9.1c). We suspect that this involves the identifiability problem that affects both PGLMM and MCMCglmm. When no phylogenetic signal exists (zero covariances in Y among taxa), the parameters b_0 and σ^2 (σ_s^2 for MCMCglmm) are confounded by the identifiability problem; no

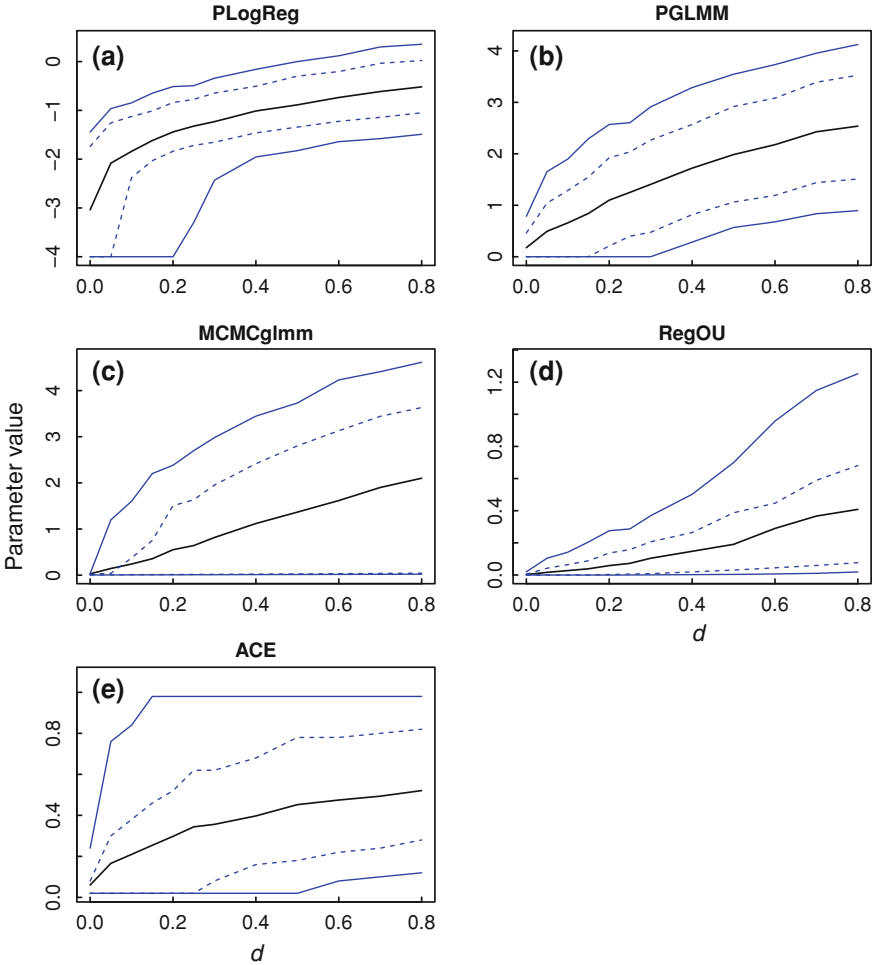

Fig. 9.1 Estimates of phylogenetic signal from **a** PLogReg, **b** PGLMM, **c** MCMCglmm, **d** RegOU, and **e** ACE. In all cases, the horizontal axis gives d of the OU branch-length transform used to general different relative branch lengths for phylogenetic trees up which the binary dependent variable evolved. The vertical axis gives the phylogenetic signal parameter for each of the five methods. In each panel, the central line is the mean parameter values, and *dashed* and *solid lines* give the 66 and 90 % inclusion intervals from 1,000 simulations at each value of d with $a = 0$. Phylogenetic trees were all assumed to be symmetrical with 64 taxa. The lower boundary of the 90 % inclusion interval corresponds to 5 % of the simulated data having estimates of no phylogenetic signal, which gives an indication of the probability that the point estimate of phylogenetic signal from each model is greater than zero. These simulations show the relative performance of the methods at identifying phylogenetic signal, with PLogReg outperforming the other methods, and able to detect signal when the value of d equals or exceeds 0.2. Note that d can exceed unity, but simulations with d set at values > 0.8 were deemed unnecessary

single pair of values gives the best model fit. The estimation approach of PGLMM uses an iterative process alternating conditional likelihoods for b_0 given σ^2 and for σ^2 given b_0, and we suspect that this acts to separate (i.e., decrease the correlation) of the estimates of these parameters. By decreasing this correlation, PGLMM will reduce the variability in the estimate of σ^2. Consistent with this explanation, for simulations with $d = 0.5$, the correlation between the estimates of $|b_0|$ and σ for PGLMM was -0.26, while for MCMCglmm the correlation between $|b_0|$ and σ_s^2 was -0.35. At $d = 0.5$, zero is well outside the 90 % inclusion interval for PGLMM while still within the inclusion interval for MCMCglmm.

RegOU did not perform well. This is presumably because it did a poor job capturing the phylogenetic covariances among values of Y. By ignoring the binary nature of the data, RegOU does not incorporate the constraints on the variances and covariances that are imposed by the nature of binary data. Specifically, the covariances are bounded by the variances in Y, and the variances in turn are set by $p(1 - p)$, which has a maximum value of $p = 0.5$. Therefore, if by chance there is a relatively large number of ones (or zeros) in a data set, then the estimate of p will be greater (or less) than 0.5 and the covariances in the data will necessarily be reduced. RegOU does not account for this, and we suspect that this is the reason for its poor power. In contrast, PLogReg explicitly accounts for the dependency of the variances and covariances on p, in effect increasing the weight of the covariances in the data to compensate for changes in p.

The performance of ACE was third after PGLMM. We expected ACE to perform better. The performance of ACE could be limited by its use of ML, in contrast to penalized likelihood used by PLogReg and REML used by PGLMM, although other explanations are possible, including the mismatch between the model underlying ACE and the simulation model that we used.

9.3.2 Regression

We investigated the performance of five methods—PLogReg, PGLMM, MCMCglmm, RegOU, and Logistf—to estimate b_1 giving the relationship between Y and the independent variable x. We simulated 1,000 data sets with $b_0 = 0$ and $b_1 = 1$ for 64 species on a symmetrical phylogenetic tree subjected to an OU branch-length transform with values of d between 0 and 2 in increments of 0.25. The values of x were assumed to have evolved under a BM model up the phylogenetic tree with no branch-length transform (i.e., $d = 1$). For each data set, we then estimated b_1 using all five methods and plotted the means from the 1,000 simulations, as well as the 66 and 90 % inclusion intervals (Fig. 9.2).

Because of differences between the simulation and statistical models, we did not expect the estimates of b_1 to be exactly 1. Nonetheless, changes in the estimates of b_1 with d used to simulate the data indicate that phylogenetic signal introduces bias in the estimates. All methods except RegOU showed upward bias

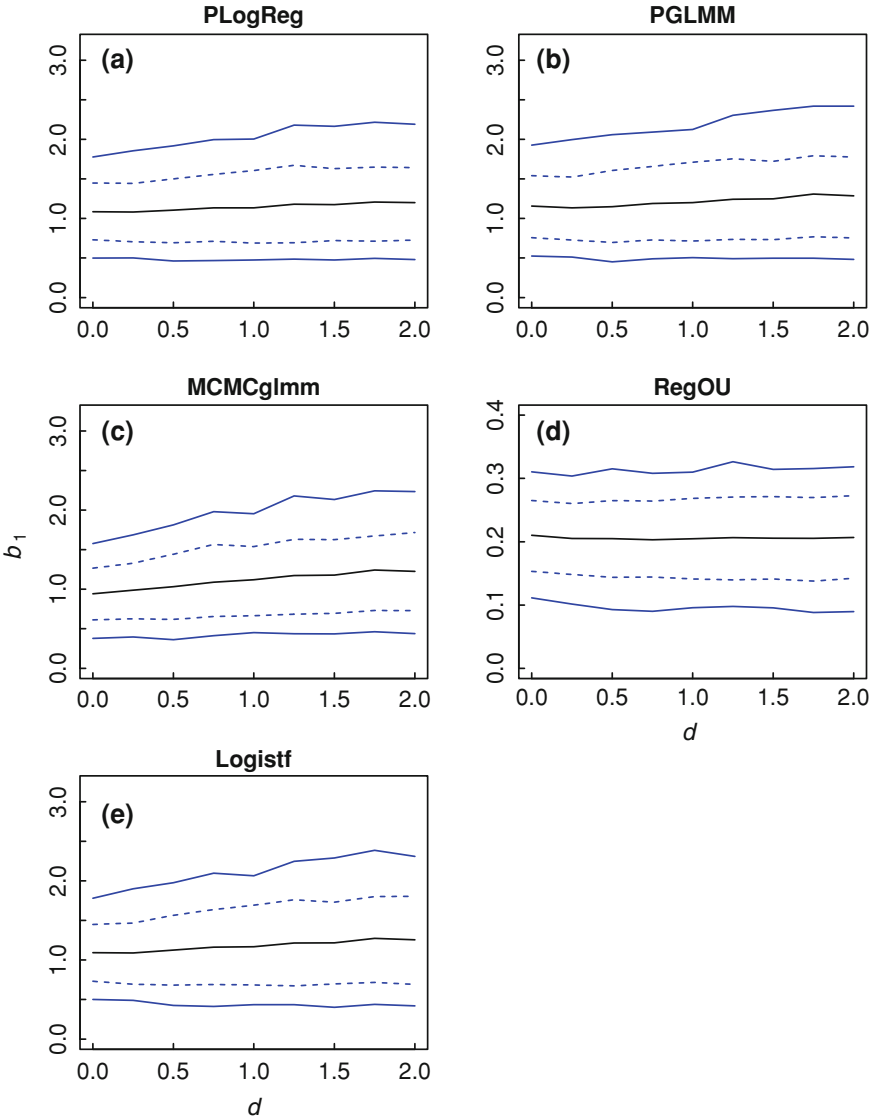

Fig. 9.2 Estimates of regression parameter b_1 from **a** PLogReg, **b** PGLMM, **c** MCMCglmm, **d** RegOU, and **e** Logistf. In all cases, the horizontal axis gives d of the OU branch-length transform used to generate different relative branch lengths for phylogenetic trees up which the binary dependent variable evolved, and in the simulations $b_1 = 1$. In each panel, the central line is the mean parameter values, and *dashed* and *solid lines* give the 66 and 90 % inclusion intervals from 1,000 simulations at each value of d with $a = 0$. Values for PGLMM and MCMCglmm were corrected by factors $(1 + c^2\sigma^2)^{-0.5}$ and $(1 + c^2\sigma_s^2 + c^2\sigma_u^2)^{-0.5}$, respectively (see text). For simulations, the independent variable x was assumed to evolve as a Brownian motion process along the specified tree. Phylogenetic trees were all assumed to be symmetrical with 64 taxa. These results show both increased bias (mean parameter values) and decreased precision (width of inclusion intervals) with increasing phylogenetic signal d

with increasing phylogenetic signal, with PLogReg showing the least (12 % when $d = 2$ vs. $d = 0$) and MCMCglmm showing the most (32 % when $d = 2$ vs. $d = 0$). Simultaneously, increasing d made the estimates of b_1 more variable for all methods, as shown by the broadening of the inclusion intervals. This loss of precision was least for RegOU (15 %) followed by PLogReg (34 %) and was greatest for Logistf (48 %). Thus, failing to account for phylogenetic signal in Logistf led to the greatest loss in precision as phylogenetic signal increased.

One of the greatest statistical concerns driving phylogenetic comparative methods is Type I errors, rejecting the null hypothesis when it is true (e.g., see Grafen 1989; Martins and Garland 1991; Garland et al. 1993; Diaz-Uriarte and Garland 1996). To investigate Type I errors, we simulated data under the null hypothesis of $b_1 = 0$ and scored for each statistical method the proportion of simulations for which the null hypothesis was rejected at the $\alpha = 0.05$ significance level (Fig. 9.3). The data were simulated up balanced, ultrametric phylogenetic trees for $N = 16$, 32, 64, 128, and 256 terminal taxa assuming substantial phylogenetic signal ($d = 1$). Values of the independent variable x were simulated either with phylogenetic signal given by BM evolution (Fig. 9.3a) or without phylogenetic signal (Fig. 9.3b). The statistical tests for H_0:$b_1 = 0$ were performed using asymptotic approximations with PLogReg, PGLMM, RegOU, and Logistf. For MCMCglmm, to give comparable information to hypothesis testing, we used the 95 % credible interval for b_1, scoring the proportion of simulations for which zero fell outside this interval.

Ideally, all methods would reject H_0:$b_1 = 0$ in 5 % of the simulated data sets when the value of b_1 in the simulations is in fact zero. Because we only performed 1,000 simulations, there will be some variability around this 5 % expectation. Specifically, for 1,000 simulations, the 95 % confidence interval for expected number of simulated data sets rejecting H_0:$b_1 = 0$ when the rejection rate is correct can be calculated to be (3.6, 6.4 %); consistent departures from this range imply incorrect Type I error rates. When phylogenetic signal exists in the independent variable x, Logistf rejects H_0:$b_1 = 0$ much more frequently, indicating an inflated Type I error rate (Fig. 9.3a). In the absence of phylogenetic signal in x, however, Logistf does not perform badly (Fig. 9.3b). This type of behavior—sensitivity of a statistical regression method to the distribution of the independent variable—is also found in phylogenetic regression with continuous dependent variables (Revell 2010). In standard regression, when phylogenetic signal is absent in the residual variation, the distribution of the independent variables does not affect the estimates and statistical tests of the corresponding regression coefficients. However, with phylogenetic signal in the residuals, the presence or absence of phylogenetic signal in the independent variables does matter. The results for Logistf show that when an independent variable is correlated with the residuals, the data contain less information about the parameters. Logistf ignores this loss of information, thereby giving incorrectly strong statistical tests leading to badly inflated Type I error rates.

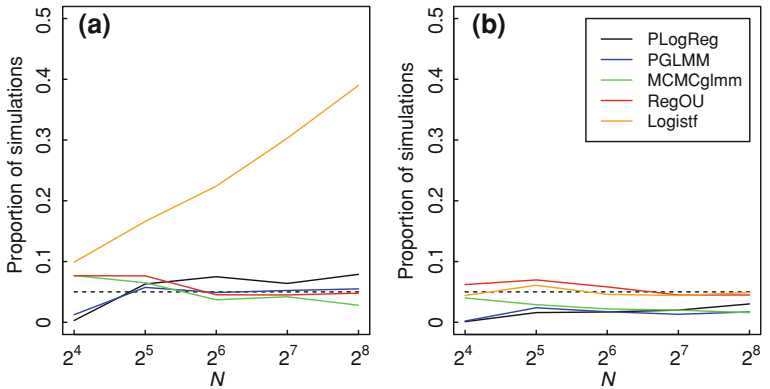

Fig. 9.3 Type I error rates (proportions of simulations in which H_0: $b_1 = 0$ was rejected when it was true) at $\alpha = 0.05$ as a function of the number of taxa in the sample (16, 32, 64, 128, and 256) on a symmetrical phylogenetic tree for PLogReg (*black*), PGLMM using approximate standard error (*blue*), MCMCglmm (*green*), RegOU (*red*), and Logistf (*orange*). *Dotted black line* indicates nominal Type I error rate of 5 %. In **a** the independent variable x is assumed to show Brownian motion phylogenetic signal, and in **b** x is assumed to have no phylogenetic signal (evolution up a star phylogeny). In both **a** and **b** there is phylogenetic signal in the residual variation given by $d = 1$ and $a = 0$. For all methods, the lines give the proportion of 1,000 simulations in which the null hypothesis $b_1 = 0$ is rejected at the $\alpha = 0.05$ level based on the approximate asymptotic distribution of b_1, except for MCMCglmm which gives the proportion of simulations in which the 95 % credible interval excludes $b_1 = 0$. Simulated data sets with Y taking few values of 0 or 1 ($\leq$1/8 values) were excluded, because these data sets will give little information for statistical fitting and a practitioner probably should not try to analyze them statistically (e.g., see Diaz-Uriarte and Garland 1996, p. 45). Results in this figure show that all methods other than PGLMM (*blue line*) show inflated Type I error rates for some values of N when there is phylogenetic signal in x (panel **a**)

The phylogenetic methods performed much better than Logistf when phylogenetic signal is present in x (Fig. 9.3a). For small sample sizes ($N = 16$), both PLogReg and PGLMM underestimated the nominal number of simulations with "significant" values of b_1, although they performed well for $N \geq 32$; this underestimate of significance (lower-than-appropriate Type I error rates) when $N = 16$ is conservative, in the sense that it means that false positives are less likely. MCMCglmm had inflated Type I error rates for $N = 16$, which is an issue of concern, although MCMCglmm performed similarly to PLogReg and PGLMM for $N \geq 32$. Overall, RegOU performed well although, like MCMCglmm, it showed inflated Type I error for $N = 16$. Without phylogenetic signal in x (Fig. 9.3b), PLogReg, PGLMM, and MCMCglmm all showed lower-than-appropriate Type I error rates, especially PLogReg and PGLMM when $N = 16$. Perhaps most surprisingly, the two best performers were RegOU and Logistf that ignore, respectively, the binary nature of the data and phylogenetic signal. These results illustrate that, in statistics, it is sometimes possible for a method to be right even when it does not incorporate all of the characteristics of the data.

To investigate statistical power, we generated curves of the probability of rejecting the null hypothesis H_0:$b_1 = 0$ with values of b_1 ranging from 0 to 1 (Fig. 9.4). The data were simulated for 64 species assuming phylogenetic signal in Y ($d = 1$). Values of the independent variable x were simulated either with phylogenetic signal given by BM evolution (Fig. 9.4a) or without phylogenetic signal (Fig. 9.4b). For PGLMM, in addition to performing the analyses using the asymptotic approximation for standard errors of b_1, we also performed a bootstrap test of H_0:$b_1 = 0$ using the 95 % bootstrap confidence intervals of b_1. To make this feasible, we used only 200 bootstrapped data sets for each of 200 simulated initial data sets, which still required 40,000 estimations for each of the 11 values of d. We did not perform a similar test of the bootstrap for PLogReg, because the computer time to run the 440,000 estimations was prohibitive. For single, real data sets, however, bootstraps can be run for PLogReg.

RegOU performed well, with a correct Type I error rate (at $d = 0$) and good power, especially when phylogenetic signal is absent in x (Fig. 9.4b). With phylogenetic signal in x (Fig. 9.4a), MCMCglmm had greatest power, at least for this sample size of $N = 64$ (but see Fig. 9.3a). In part, this might be due to the greater upward bias shown by MCMCglmm (Fig. 9.2c), which makes it more likely for the 2.5 % tail of the distribution of the estimate of b_1 to lie above zero. Using the asymptotic approximations for the standard error of b_1, both PLogReg and PGLMM had slightly lower power than MCMCglmm and comparable power to RegOU. The bootstrap results for PGLMM had the correct Type I error rates at $d = 0$; correct Type I error rates are guaranteed with bootstrapping, provided the data are well fit by the PGLMM model, because bootstrapping is based on simulating data and scoring those for which H_0:$b_1 = 0$ is rejected. Despite having the correct Type I error rates when $d = 0$, the PGLMM bootstrap had lower power than the other methods. Despite this low power, because the bootstrapping guarantees lack of Type I errors, it provides the most "secure" results.

Overall, the power curves show that all methods have roughly similar power. However, given the inflated Type I error rates shown by all methods except for PGLMM for some sample sizes (Fig. 9.3a), all phylogenetic methods run the risk of giving false positive results. Generally, Type I errors are of greater statistical concern than Type II errors (accepting the null hypothesis when it is false), and the Type I errors shown by the simulations should lead to caution when interpreting results. Indeed, many statisticians would not move on to consider statistical power unless Type I error rates were assured at a nominal level, such as $\alpha = 0.05$ (e.g., Martins and Garland 1991).

9.4 Method Comparison with Real Data

To compare methods using real data, we used an example provided by Brashares et al. (2000) for 75 species of African antelope; this data set was also analyzed using PLogReg in Ives and Garland (2010). We tested the hypothesis proposed by

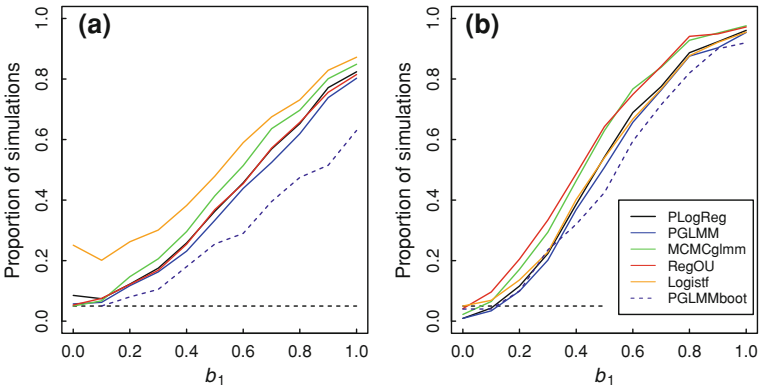

Fig. 9.4 Power curves for estimates of the regression parameter b_1 from PLogReg (*black*), PGLMM using approximate standard error (*blue*), PGLMM using bootstrapped confidence intervals (*dashed dark blue*), MCMCglmm (*green*), RegOU (*red*), and Logistf (*orange*). The power curves give the proportion of simulations in which the null hypothesis H_0: $b_1 = 0$ was rejected at the $\alpha = 0.05$ level for 64 taxa evolving on a symmetrical phylogenetic tree. In **a** the independent variable x is assumed to show Brownian motion phylogenetic signal, and in **b** x is assumed to have no phylogenetic signal (evolution up a star phylogeny). In both **a** and **b**, there is phylogenetic signal in the residual variation given by $d = 1$ and $a = 0$. For all methods, the lines give the proportion of 1,000 simulations in which the null hypothesis H_0: $b_1 = 0$ was rejected at the $\alpha = 0.05$ level, except for MCMCglmm which gives the proportion of simulations in which the 95 % credible interval excludes $b_1 = 0$, and the bootstrapped results for PGLMM that are based on 200 simulations. These results show that all methods have similar power, with the bootstrapped power curve for PGLMM having the lowest power

Jarman (1974) that species living in larger groups are more likely to flee or fight predators ($Y = 1$), whereas solitary or pair-living species are more likely to hide ($Y = 0$). Group size ranges between 1 and 70, and we treated $\log_{10}$-transformed group size as a continuous variable. Because body size is likely also to affect antipredator behavior, with larger-bodied species more likely to flee/fight than hide, we followed Brashares et al. (2000) and also included $\log_{10}$ body mass as a second independent, continuously valued variable. To stabilize the statical analyses, both independent variables were standardized to have mean equal to zero and standard deviation equal to one; this also makes the regression coefficients represent effect sizes of the independent variables whose magnitudes reflect the size of effect of the variable (as is traditionally done in path analysis, Chap. 8). Group size had phylogenetic signal (RegOU: $d = 0.84$, conf. interval = (0.52, 1.07); Lavin et al. 2008), as did body size ($d = 0.99$, conf. interval = (0.77, 1.20)), suggesting that phylogenetic signal in Y will present statistical challenges. Here, we summarize the results of the analyses, and in the online practical material (http://www.mpcm-evolution.org), we present this example as a tutorial.

All methods revealed a strong positive effect of group size on the response of antelope to predators (Table 9.1, Fig. 9.5); antelopes with larger group sizes were more likely to flee/fight than to hide. Because the methods incorporate the regression

Table 9.1 Comparison among five methods of estimating regression coefficients for the effects of $\log_{10}$ group size and $\log_{10}$ body mass on the antipredator behavior (0 = hide, 1 = flee or fight) of 75 antelope species

Parameter[a]	Estimate	Approx. SE[b]	95 % confidence/ credible interval[c]	p-value	Bootstrap mean[d]	95 % bootstrap confidence interval
PLogReg						
a	0.50				−0.49	(−4, 4)
b_0	−0.82	0.87	(−2.54, 0.90)	0.34	−0.86	(−3.67, 1.53)
b_1 (body mass)	0.096	0.45	(−0.80, 0.99)	0.84	0.13	(−1.08, 1.34)
b_2 (group size)	1.36	0.49	(0.39, 2.33)	0.007	1.71	(0.47, 3.66)
PGLMM						
σ^2	7.16				2.86	(0.00, 8.0)
b_0	−0.70	0.79	(−2.27, 0.87)	0.38	−1.13	(−3.27, 0.35)
b_1 (body mass)	0.22	0.60	(−0.98, 1.42)	0.71	0.51	(−0.98, 2.20)
b_2 (group size)	1.45	0.66	(0.13, 2.77)	0.031	1.66	(0.56, 3.24)
MCMCglmm						
σ_s^2	4.28		(1.50, 13.9)		0.83	(0.01, 3.28)
σ_u^2	1				1	
b_0	−5.23		(−13.2, −1.12)		−3.92	(−10.1, −0.34)
b_1 (body mass)	0.54		(−2.56, 4.03)		0.49	(−1.81, 3.23)
b_2 (group size)	4.20		(1.50, 13.9)		3.07	(0.68, 6.55)
RegOU						
d	0.62					
b_0	0.39	0.20	(−0.008 0.79)	0.47		
b_1 (body mass)	0.16	0.070	(0.021, 0.30)	0.026		
b_2 (group size)	0.16	0.056	(0.049, 0.27)	0.007		
Logistf						
b_0	−1.57	0.96	(−3.47, 0.33)	0.09	−1.59	(−3.66, 0.18)
b_1 (body mass)	−1.17	0.87	(−2.89, 0.55)	0.16	−1.16	(−3.01, 0.39)
b_2 (group size)	4.57	1.25	(2.10, 7.05)	<0.0001	4.56	(2.27, 7.44)

[a] All independent variables were standardized to have mean 0 and variance 1 prior to analysis
[b] Standard errors of the estimates and confidence intervals were obtained using the asymptotic approximations
[c] Approximate confidence intervals for the frequentist methods were computed using asymptotic approximations. The credible intervals for MCMCglmm were computed in the fitting process
[d] Parametric bootstrapping was performed by simulating 1,000 data sets to obtain means and confidence intervals

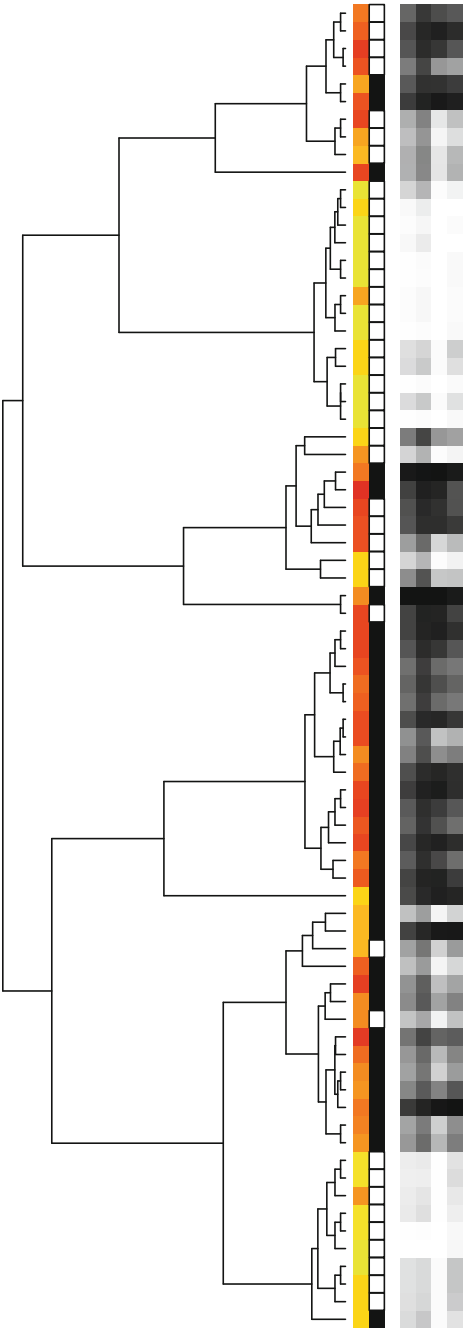

Fig. 9.5 Analyses of antipredator behavior (0 = hide, 1 = flee or fight) of 75 antelope species as it depends on $\log_{10}$ group size (see results shown in Table 9.1). The *first column* to the *right* of the phylogenetic tree gives $\log_{10}$ group size, a continuously valued independent variable, colored from small (*yellow*) to large (*red*). The second column gives the trait value Y (hide = white, flee/fight = black). The remaining 4 columns give the fit of the models—PLogReg, PGLMM, MCMCglmm, and Logistf—as the probability of fleeing/fighting predicted by the models (Table 9.1) scaled from fleeing/fighting with probability zero (*white*) to one (*black*). These fits also incorporate the information from the second independent variable, $\log_{10}$ body size. All methods revealed the effect of group size on the response variable (Table 9.1), and their similar predictions are shown by similar patterns in the last 4 columns

coefficients differently, their values cannot be compared directly. Nonetheless, all methods showed significant departure of b_2 from zero. Although this main conclusion would be reached regardless of the method used, the results nonetheless show properties of the methods that we will discuss for each method in turn.

For PLogReg, bootstrapping shows that the value of b_2 is an overestimate, because the bootstrap mean is 1.71 compared to the value of 1.36 from the data that were used to parameterize the simulation bootstraps. Furthermore, the bootstrap confidence intervals are wider than those obtained from the approximate standard errors (approximate (0.39, 2.33); bootstrapped (0.47, 3.66)). Simultaneously, the bootstrap confidence interval for a ranging from the minimum to maximum values (−4 to 4) indicates that PLogReg is unable to determine whether there is phylogenetic signal in the residuals.

Like PLogReg, the bootstrap of PGLMM shows upward bias in the estimate of b_2. The bootstrap confidence intervals show no statistically significant phylogenetic signal in the residuals (lower bound for σ^2 is zero). Nonetheless, due to the upward bias in the estimates, the lower bound of the confidence interval could also be upward biased. To test for this, we also performed a bootstrap under the null hypothesis H_0:$b_2 = 0$ by fitting the model without group size x_2, simulating data sets from the fitted model, and then fitting the full model including group size to the simulated data sets. Only 1/1000 simulations had an estimated value of $b_2 > 1.45$ (the value observed in the real data set), implying strong rejection of the null hypothesis H_0:$b_2 = 0$.

Despite using essentially the same statistical models, the parameter estimates from MCMCglmm were substantially different from PGLMM. We produced "bootstrap confidence intervals" for MCMCglmm by simulating 1,000 data sets from the fitted MCMCglmm model and reporting the range covered by 95 % of the estimated parameters. This procedure showed downward bias in the estimate of b_2 although little bias in b_1.

RegOU showed not only a statistically significant effect of group size but also an effect of body size. Based on the simulations that show RegOU gives correct Type I errors and has good power, we are inclined to trust these results with respect to hypothesis testing. Nonetheless, it is not possible to interpret the values of the parameter estimates in a meaningful way; they cannot be converted into the probability that predators flee or fight. This also means that bootstrapping is impossible, leading us to rely on the simulations (Figs. 9.2, 9.3 and 9.4) to give us faith in the results.

Logistf gave, on the face of it, very good statistical properties. The effect of group size was highly significant, and the bootstrapping showed that there is little bias and that the approximate confidence intervals are accurate. However, because both group size and body size show phylogenetic signal, we know that Logistf is likely to suffer from severely inflated Type I errors (Fig. 9.3). Therefore, we cannot trust the statistical results. This shows that even "good" superficial statistical behavior can be underlain by serious statistical mistakes.

Assessing the methods by their fits to the data, all models gave similar predictions for the probability that antelope trait flee/fight (Fig. 9.5). There are clear

cases in which all models predict $Y = 1$ with high probability, yet in fact $Y = 0$ (compare column 2 with columns 3–6). These mis-predictions also appear to have a phylogenetic pattern, with closely related species showing $Y = 0$ despite predictions otherwise. Nonetheless, none of the methods identified statistically significant phylogenetic signal in the residual variation (Table 9.1).

9.5 Discussion

Of the several methods now available for analyzing phylogenetic binary data, our comparisons gave no single "winner" that performed best under all situations and for all questions asked about the data. If we had considered a greater range of phylogenetic trees, sample sizes, and data characteristics (e.g., including multiple regressions), then a "winner" might have been even less apparent. This is not surprising, given the complexity of these methods and of the phylogenetically evolved data they are trying to analyze. Below, we try to give simple guidelines for different situations that might commonly arise, treating first the detection of phylogenetic signal and then regression with independent variables. We emphasize, however, that multiple methods should be tried for any data set; data are generally far more time-consuming to generate than are statistical models to run, so once data are in hand, it makes sense to run multiple tests. If they all give the same results, this is reassuring. If they do not, then our analyses have hopefully pointed to reasons for the differences and helped identify which methods to trust. Nonetheless, we stress that our simulations are far from exhaustive, and it is best to tailor simulations to particular data sets. Moreover, we hope and expect new methods will be developed which can be applied to old data, as raw data are now becoming routinely deposited through online supplementary material and communal repositories.

9.5.1 Phylogenetic Signal (Regression Without Independent Variables)

At least in our simulations, PLogReg was the most powerful method for detecting phylogenetic signal in the absence of independent variables. It was the most likely to show the existence of phylogenetic signal when it was in fact present (Fig. 9.1). On the downside, PLogReg is also the most computer intensive of the methods considered here. For large numbers of taxa (e.g., $N = 256$), bootstrapping can take a day on a typical personal computer. Speed might be considerably helped, however, by the methods of Ho and Ane implemented in **phylolm** {R} (Ho and Ane 2014).

In some situations, other methods are preferable. Following PLogReg, PGLMM performs well and typically runs 100–1,000-fold faster. Therefore, for large data sets or for extensive bootstrapping, PGLMM might be preferred. MCMCglmm, as

a Bayesian method, gives information about the distributions and correlations among parameter estimates, and MCMCglmm as a package gives nice tools that make it possible to visualize and extract this information. This information is helpful for diagnosing problems with model estimation; it can be used in a way similar to bootstrapping to approximate the joint distribution of parameter estimates. An advantage of MCMCglmm is that rather than iterating the estimation procedure many times as required for bootstrapping, MCMCglmm provides diagnostic information in a single estimation, making it functionally very fast.

Unlike the other methods, ACE provides estimates of ancestral states and uncertainty in these estimates. This can be useful to illustrate possible evolutionary sequences of events leading to the current distribution of traits among taxa. Personally, we are cautious about trying to infer much about evolutionary history from information on only extant species (Garland et al. 1999; Bonine et al. 2005; Chap. 22), yet mapping-inferred ancestral trait changes onto phylogenies can be useful in conjunction with historical information about possible drivers of evolution, such as environmental changes that cause alterations in the selective regime (Hansen and Orzack 2005).

9.5.2 Regression

When an analysis includes independent variables, the main goals will likely be to test the statistical significance and characterize the relationship between the dependent and independent variables, that is, estimate the regression coefficients and associated confidence intervals or statistical tests. Guidelines for method selection for regression analyses are more complicated than those for detecting phylogenetic signal, so we discuss different scenarios separately.

9.5.2.1 Independent Variables Lack Phylogenetic Signal

When independent variables show no phylogenetic signal, all methods worked surprisingly well. In fact, there are few statistical problems introduced by phylogenetic signal in the residuals; even Logistf worked well, with little sign of inflated Type I errors. Therefore, it makes sense, before performing any other analysis, to first test for phylogenetic signal in the independent variables. If the estimates of phylogenetic signal are zero, then try Logistf or another logistic regression package (but use the Firth correction). It is necessary to recognize that by "no phylogenetic signal," we mean that the estimate of phylogenetic signal is exactly zero; absence of statistically significant phylogenetic signal is probably not a strict enough requirement to ignore phylogenetic signal all together, because the power of statistical tests for signal is often low (Blomberg et al. 2003).

9.5.2.2 Testing for Significant Regression Coefficients

If the goal of the analysis is to test the null hypothesis H_0: $b_1 = 0$, then RegOU or similar methods that ignore the binary nature of the dependent variable give a simple approach that in our simulations was remarkably statistically robust. This is not the heresy it might initially seem. The central limit theorem is remarkable: Given enough samples, the sum of their values will approach a Gaussian distribution, regardless of the distribution of any single sample. In principle, ignoring the unavoidable variation caused by Y taking only values of zero and one, RegOU should suffer loss of power compared to methods that account for this variation, yet we did not see a large loss of power in simulations with $N = 64$ (Fig. 9.4). A major disadvantage of RegOU, however, is that it does not give a model that fits the data. Therefore, it is hard to interpret the meaning of the regression coefficients, and it is not possible to simulate data from the model that has the statistical properties of the original data. This makes parametric bootstrapping impossible.

9.5.2.3 Fitting a Binary Regression Model

PLogReg, PGLMM, and MCMCglmm all performed reasonable well in fitting a model to data, although all had worrying defects. In particular, using the asymptotic approximations to the standard errors of the parameters, PLogReg tended to give inflated Type I errors for larger samples sizes, whereas MCMCglmm gave inflated Type I errors for smaller samples. Furthermore, all three methods showed lower-than-expected Type I errors under some situations, implying loss of statistical power (Fig. 9.3).

A solution to the Type I errors for PLogReg and PGLMM is to bootstrap the statistical tests for H_0:$b_1 = 0$. Bootstrapping guarantees against inflated Type I errors. Furthermore, bootstrapping the estimates of b_1 will give the most accurate confidence intervals and also show possible bias in the parameter estimates. Similarly, investigation of the joint distribution of parameter estimates in MCMCglmm will show possible problems with parameter estimation. Thus, although these methods have potential problems, those problems can be identified, and bootstrapping for PLogReg and PGLMM can solve the problem of inflated Type I errors. These two methods, used with parametric bootstrapping, provide the only approaches to fitting an appropriate model to phylogenetic binary data that have yet been validated.

9.6 Summary and Future Directions

Analyzing binary comparative data while accounting for potential phylogenetic correlations is not straightforward, but we hope our rough guidelines will help. It is important at the onset of analyses, however, to carefully consider your data. If you

have, for example, $N = 16$ taxa of which only three show one of the two possible trait states, then you should probably not embark on an analysis at all, particularly if the three taxa are closely related; the amount of information available in the data is unlikely to yield trustworthy statistical results. Before analyzing the dependent variable, test for phylogenetic signal in the independent variables; strong phylogenetic signal should suggest being particularly cautious in a regression analysis. Similarly, check for collinearity among the independent variables and remove those that seem unlikely to be informative on a priori grounds. Visually inspect the data: Are there observable patterns in the dependent variable? Do model fits look consistent with the data (e.g., Fig. 9.5)? If a statistical test tells you that a pattern is significant, then you should be able to see it in the data. After initial analyses, perform diagnostics by bootstrapping or by examining the parameter distributions from MCMCglmm. Use multiple methods, hoping that they give similar results; if they do not, then try to figure out why and which (if any) to trust. Finally, do not hesitate to use a model that you know is not strictly appropriate—phylogenetic regression for continuous traits, such as RegOU—as it still might give results that are statistically robust for hypothesis testing.

Phylogenetic models for the analysis of binary dependence variables need further statistical development. A simple, fast, robust method that does not suffer inflated Type I errors is still not available. Improvements in algorithms (e.g., **phylolm** {R}) will help. Another possible avenue is to pursue probit models; for MCMCglmm, the probit model had better statistical properties than the logit model, although probit models are not available for frequentist methods. A limitation of the three available methods we investigated is the absence of a true ML or REML score that can be used for model selection. Practically, this is not a huge hindrance, because all of the models can be used with classical backward or forward stepwise regression using, for example, P-values to decide which terms to include. Nonetheless, the popularity of backward and forward selection procedures has waned, and most researchers will now want an AIC score or a related information-theoretic metric. Finally, we have only investigated methods for binary data, but the same issues (although probably less severe) will appear for Poisson and other non-Gaussian data. Lots of problems are yet to be solved.

References

Blomberg SP, Garland T Jr, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24(3):127–135. doi:10.1016/j.tree.2008.10.008

Bonine KE, Gleeson TT, Garland T Jr (2005) Muscle fibre-type variation in lizards (Squamata) and phylogenetic reconstruction of hypothesized ancestral states. J Exp Biol 208:4529–4547

Brashares JS, Garland T Jr, Arcese P (2000) Phylogenetic analysis of coadaptation in behavior, diet, and body size in the African antelope. Behav Ecol 11(4):452–463

Diaz-Uriarte R, Garland T Jr (1996) Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Syst Biol 45(1):27–47

Diggle P, Heagerty P, Liang K, Zeger S (2004) Analysis of longitudinal data, 2nd edn. Oxford University Press, Oxford

Dlugosz EM, Chappell MA, Meek TH, Szafrańska P, Zub K, Konarzewski M, Jones JH, Bicudo JEPW, Careau V, Garland T Jr (2013) Phylogenetic analysis of mammalian maximal oxygen consumption during exercise. J Exp Biol 216:4712–4721

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

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

Felsenstein J (1988) Phylogenies and quantitative characters. Annu Rev Ecol Syst 19:445–471

Felsenstein J (2012) A comparative method for both discrete and continuous characters using the threshold model. Am Nat 179(2):145–156. doi:10.1086/663681

Firth D (1993) Bias reduction of maximum likelihood estimates. Biometrika 80(1):27–38

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat 160:712–726

Garland T Jr, Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer-simulation. Syst Biol 42(3):265–292

Garland T Jr, Harvey PH, Ives AR (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst Biol 41:18–32

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

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis, 1st edn. Chapman and Hall, London

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

Grafen A (1989) The phylogenetic regression. Trans R Soc Lond B, Biol Sci 326:119–157

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

Hadfield JD (2012) MCMCglmm course notes. http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf

Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23(3):494–508. doi:10.1111/j.1420-9101.2009.01915.x

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351

Hansen TF, Orzack SH (2005) Assessing current adaptive and phylogenetic inertia explanations of trait evolution: the need for controlled comparisons. Evolution 59:2063–2072

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford

Heinze G, Ploner M, Dunkler D, Southworth H (2013) logistf: Firth's bias reduced logistic regression. Vol R package version 1.21

Heinze G, Schemper M (2002) A solution to the problem of separation in logistic regression. Stat Med 21:2409–2419

Ho LST, Ane C (2014) A linear-time algorithm for gaussian and non-gaussian trait evolution models. Systematic Biology in press

Ives AR, Garland T (2010) Phylogenetic logistic regression for binary dependent variables. Syst Biol 59(1):9–26. doi:10.1093/sysbio/syp074

Ives AR, Helmus MR (2011) Generalized linear mixed models for phylogenetic analyses of community structure. Ecol Monogr 81:511–525

Jarman PJ (1974) The social organisation of antelope in relation to their ecology. Behaviour 48:215–267

Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee T-C (1985) The theory and practice of econometrics, 2nd edn. Wiley, New York

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464

Lavin SR, Karasov WH, Ives AR, Middleton KM, Garland T Jr (2008) Morphometrics of the avian small intestine, compared with non-flying mammals: a phylogenetic approach. Physiol Biochem Zool 81:526–550

Martins EP, Diniz JAF, Housworth EA (2002) Adaptive constraints and the phylogenetic comparative method: a computer simulation test. Evolution 56(1):1–13

Martins EP, Garland T Jr (1991) Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557

Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am Nat 149:646–667. Erratum 153:448

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London

McCulloch CE, Searle SR, Neuhaus JM (2008) Generalized, linear, and mixed models. Wiley, Hoboken, NJ

Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc Lond Ser B Biol Sci 255(1342):37–45

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

Revell LJ (2010) Phylogenetic signal and linear regression on species data. Methods Ecol Evol 1:319–329

Revell LJ (2012) Analyzing continuous character evolution on a phylogeny. Integr Comp Biol 52:E145–E145

Revell LJ, Harmon LJ, Collar DC (2008) Phylogenetic signal, evolutionary process, and rate. Syst Biol 57(4):591–601. doi:10.1080/10635150802302427

Rezende EL, Diniz JAF (2012) Phylogenetic analyses: comparing species to infer adaptations and physiological mechanisms. Compr Physiol 2(1):639–674. doi:10.1002/cphy.c100079

Schluter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711

Villemereuil P, Gimenez O, Doligez B (2013) Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. Methods Ecol Evol 4(3):260–275. doi:10.1111/2041-210x.12011