The LINNEAN SOCIETY of London — BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2014, **113**, 1167–1178. With 4 figures

REVIEW ARTICLE

On the use of log-transformation versus nonlinear regression for analyzing biological power laws

GARY C. PACKARD*

Department of Biology, Colorado State University, Fort Collins, CO 80523-1878, USA

Received 3 June 2014; revised 8 July 2014; accepted for publication 8 July 2014

Xiao and colleagues re-examined 471 datasets from the literature in a major study comparing two common procedures for fitting the allometric equation $y = ax^b$ to bivariate data (Xiao *et al*., 2011). One of the procedures was the traditional allometric method, whereby the model for a straight line fitted to logarithmic transformations of the original data is back-transformed to form a two-parameter power function with multiplicative, lognormal, heteroscedastic error on the arithmetic scale. The other procedure was standard nonlinear regression, whereby a two-parameter power function with additive, normal, homoscedastic error is fitted directly to untransformed data by nonlinear least squares. Xiao and colleagues articulated a simple (but explicit) protocol for fitting and comparing the alternative models, and then used the protocol to examine each of the datasets in their compilation. The traditional method was said to provide a better fit in 69% of the cases and an equivalent fit in another 15%, so the investigation appeared to validate findings from a large majority of prior studies on allometric variation. However, focus for the investigation by Xiao and colleagues was overly narrow, and statistical models apparently were not validated graphically in the scale of measurement. The present study re-examined a subset of the cases using a larger pool of candidate models and graphical validation, and discovered complexities that were overlooked in their investigation. Some datasets that appeared to be described better by the traditional method actually were unsuited for use in an allometric analysis, whereas other datasets were not described adequately by a two-parameter power function, regardless of how the model was fitted. Thus, conclusions reached by Xiao and colleagues are not well supported and their paradigm for fitting allometric equations is unreliable. Future investigations of allometric variation should adopt a more holistic approach and incorporate graphical validation on the original arithmetic scale.

ADDITIONAL KEYWORDS: allometry – scaling.

INTRODUCTION

Biologists have recently witnessed a vigorous debate about methods for fitting statistical models to bivariate data of the kind commonly used in studies of allometric variation (Kerkhoff & Enquist, 2009; Packard, 2009; Cawley & Janacek, 2010; Packard, Boardman & Birchard, 2010). The debate has centered on two common procedures. One of the procedures is the traditional allometric method, whereby a straight line with additive, normal, homoscedastic error first is fitted to logarithmic transformations and the resulting model:

$$\log(y_i) = \log(a) + b\log(x_i) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \qquad [1]$$

then is back-transformed (exponentiated) to form a two-parameter power function with multiplicative, lognormal, heteroscedastic error:

$$y_i = ax_i^{b} * \exp(\varepsilon_i) \quad \varepsilon_i \sim N(0, \sigma^2) \qquad [2]$$

on the arithmetic (linear) scale. The other method is standard nonlinear regression, whereby a model for a two-parameter power function with additive, normal, homoscedastic error:

$$y_i = ax_i^{b} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \qquad [3]$$

is fitted directly to untransformed observations by nonlinear least squares. The alternative procedures

*Current address: 865 Three Corner Gate Rd, Livermore, CO 80536-9154, USA. E-mail: gary.packard@colostate.edu

commonly yield very different estimates for coefficients in fitted models, so the debate and its outcome are important to all those who are concerned with describing and interpreting patterns of allometric variation.

The issue appeared to be resolved when four different research groups recently reported (1) that the traditional method is generally superior to standard nonlinear regression for describing allometric variation and (2) that lognormal error is more common than normal error in data for both plants and animals (Mascaro *et al*., 2011, 2014; Xiao *et al*., 2011; Ballantyne, 2013; Lai *et al*., 2013). These conclusions were based primarily on Akaike's information criterion (AIC), which was used in all four studies to discriminate between the alternative models and which typically was lower (better) for models fitted by traditional allometry than for those fitted by standard nonlinear regression. The four studies apparently validated the results of published research that made use of the traditional allometric method and implicitly (or even explicitly) encouraged investigators to continue to use the traditional method in future work on allometric variation.

However, results of the investigations are not as clear-cut as they appear to be. First, the studies were narrowly focused on a pair of two-parameter power models [Mascaro *et al*. (2011, 2014) considered a third possibility but did not pursue its implications] and questions of general importance about quality of fit were not addressed. Second, comparisons of error structure for the alternative models were equivocal because lognormality was confounded with heteroscedasticity in models fitted by the traditional procedure and normality with homoscedasticity in models fitted by standard nonlinear regression (Packard, 2014a). In other words, differences in performance that were attributed to differences in distributions for residuals (normal versus lognormal) could just as easily have been a result of differences in variance (homoscedastic versus heteroscedastic). For these several reasons, conclusions of the investigations are not well supported.

The present critique focuses on the investigation by Xiao *et al*. (2011) because it was undertaken specifically to compare the aforementioned methods for fitting allometric equations to bivariate observations and because of the extraordinary sample that was compiled for study. The investigation was based on a re-examination of 471 datasets gleaned from the literature (compared to 1, 4, and 6 datasets in the three other investigations) and, consequently, has the greatest potential to influence the conduct of future research on allometric variation. The present study re-analyzed four of the datasets compiled by Xiao *et al*. (2011: supplement 1) to illustrate complexities that were overlooked in the course of their investigation and that cast doubt on the reliability of their paradigm for fitting statistical models to bivariate data. Each of the datasets is referenced only by the identification number used by Xiao *et al*. (2011) and predictor and response variables are labelled only as x and y, respectively; these conventions also were followed by Xiao *et al*. (2011). Information on sources for the data are available in their appendix B (Xiao *et al*., 2011). The order of presentation for examples in the present study illustrates a spectrum in the quality of datasets and in the goodness of fit of models estimated by the traditional allometric method. Xiao *et al*. (2011) also performed Monte Carlo simulations, but the outcome of the simulations was equivocal because of the narrow focus on the aforementioned models for simple power functions and because of an element of circularity in the design and execution of this component of the investigation.

MATERIAL AND METHODS

The first step in each of the four case studies presented here was to display logarithmic transformations (base 10) of the original data on a bivariate plot and then evaluate the assumption of linearity that is fundamental to application of the traditional allometric method (Reeve, 1940; Kavanagh & Richards, 1942; Richards & Kavanagh, 1945). The assumption was evaluated by fitting both a straight line and a quadratic polynomial by the method of least squares and by comparing the fits by nested analysis of variance (Ritz & Streibig, 2008: 103). If the quadratic was not a significantly better fit, the assumption of linearity was judged to be satisfied. A plot of standardized residuals against fitted values was also examined to assess the assumption of homoscedasticity and to identify potential outliers.

A two-parameter power function with multiplicative, lognormal, heteroscedastic error [2] then was estimated on the original scale by back-transforming the model for the straight line fitted to logarithms [1]. Eight regression models, with different functional form and different assumptions about random variation, next were fitted directly to untransformed data, thereby creating a pool of nine candidate models for describing the distribution of the original observations. An even larger pool of candidates could have been assembled (Marshall, Bode & White, 2013), but the additional models would have been substantially more complex than the simple ones used in the present study.

Four of the regression models in the candidate pool (i.e. straight lines and power functions with and without intercepts) were fitted by standard linear and nonlinear least squares. The linear models were included because they provide a different way to fit a power function with an allometric exponent of 1 (Huxley, 1932: 241). The four models assumed additive, normal, homoscedastic error and were of the general form:

$$y_i = f(x_i, \beta) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \qquad [4]$$

where β is a vector of parameters in a linear or nonlinear function, *f*, relating the response variable, *y*, to the measure of size, *x*. Models such as these are the default output from most statistical software.

Another four regression models were fitted by generalized nonlinear least squares. These analyses assumed additive, normal, heteroscedastic error, with variance that was modeled as a power of the fitted (mean) value for the response variable:

$$\text{var}(\varepsilon_i) = \sigma^2 * (f(x_i, \beta))^{2\theta}$$

so that the models were of the general form:

$$y_i = f(x_i, \beta) + \varepsilon_i \quad \varepsilon_i \sim N\left(0, \left(\sigma^2 * (f(x_i, \beta))^{2\theta}\right)\right) \qquad [5]$$

Theta, which is an additional parameter in the fitted model (Pinheiro & Bates, 2000: 210; Ritz & Streibig, 2008: 74; Zuur *et al*., 2009: 78), determines the spread in variance as size increases. Thus, variance is typically an increasing function of the predicted value for *y*, in much the same manner that it is an increasing function of prediction in models fitted by the traditional method [2].

Standardized residuals were calculated for the homoscedastic regression models by dividing each ordinary residual by the standard error of the estimate (i.e. by the standard deviation for residuals), whereas normalized residuals were computed for the heteroscedastic regressions by dividing each ordinary residual by the square root of the variance at the corresponding level for $\hat{y}$ (Zuur *et al*., 2009: 85). Plots of these residuals against fitted values for the response were then examined for evidence of persistent (undesirable) pattern and potential outliers. Approximately 99% of residuals typically fall between −3 and +3, so more extreme values identified potential outliers. Tests for outliers and for normality (normal probability plots) were performed when needed (Packard, 2013).

All models were fitted with the nlrwr package in R, version 2.13.2 (R Core Development Team) (for simple script, see the Supporting information, Table S1). Models fitted by regression and by traditional allometry were compared by graphical analysis on the untransformed scale (Anscombe, 1973) and by AIC (Burnham & Anderson, 2002). AICs for models fitted directly to untransformed data were taken from the output from R but those for models estimated by traditional allometry had to be recalculated to accommodate the change in scale attending back-transformation of the response variable (Xiao *et al*., 2011; Ballantyne, 2013; Lai *et al*., 2013). The lowest AIC in a set of analyses identified the best fit in the pool of candidate models [but see also Packard (2013)]. The same standard that was used by Xiao *et al*. (2011) was applied: if AIC for an alternative model differed from that of the best fit by no more than 2, the models were considered to be equivalent; however, if AICs differed by more than 2, the second model was judged to have insufficient support.

RESULTS

Example 1 (dataset #6)

The bivariate distribution for the 29 observations comprising dataset #6 was linearized by transforming both *x* and *y* (Fig. 1A); outliers were not apparent (Fig. 1A, B); and the plot of residuals gave no indication of heteroscedasticity in log *y* (Fig. 1B). Linearization of the distribution was necessary, of course, because curvilinearity in logarithmic domain would violate a fundamental requirement of the traditional allometric method (Reeve, 1940; Kavanagh & Richards, 1942; Richards & Kavanagh, 1945). Inasmuch as assumptions underlying the traditional method appear to have been satisfied, the analysis was continued.

AIC indicated that the model fitted by the traditional allometric method was far and away the best model in the candidate pool (Table 1). None of the other candidate models (including the standard two-parameter power function with additive, normal, homoscedastic error) received any support whatsoever relative to that of the model estimated by traditional allometry. Thus, the equivalent analysis of this dataset by Xiao *et al*. (2011) presumably contributed to their impression that the traditional method is generally better than standard nonlinear regression for fitting an allometric equation. The implication, of course, is that the model fitted by the traditional method also describes the dominant pattern in the bivariate distribution of untransformed observations.

However, logarithmic transformation has the potential to make ugly data look remarkably good (Packard, 2011) and the current dataset provides a case in point. When untransformed observations were displayed on a bivariate plot, it was immediately apparent that the distribution is problematic: 27 data points are grouped in a narrow column at the low end of the size distribution, two data points are in column at the high end, and a substantial gap separates the two groups (Fig. 1C). No unitary pattern is apparent in the distribution. The power function estimated by

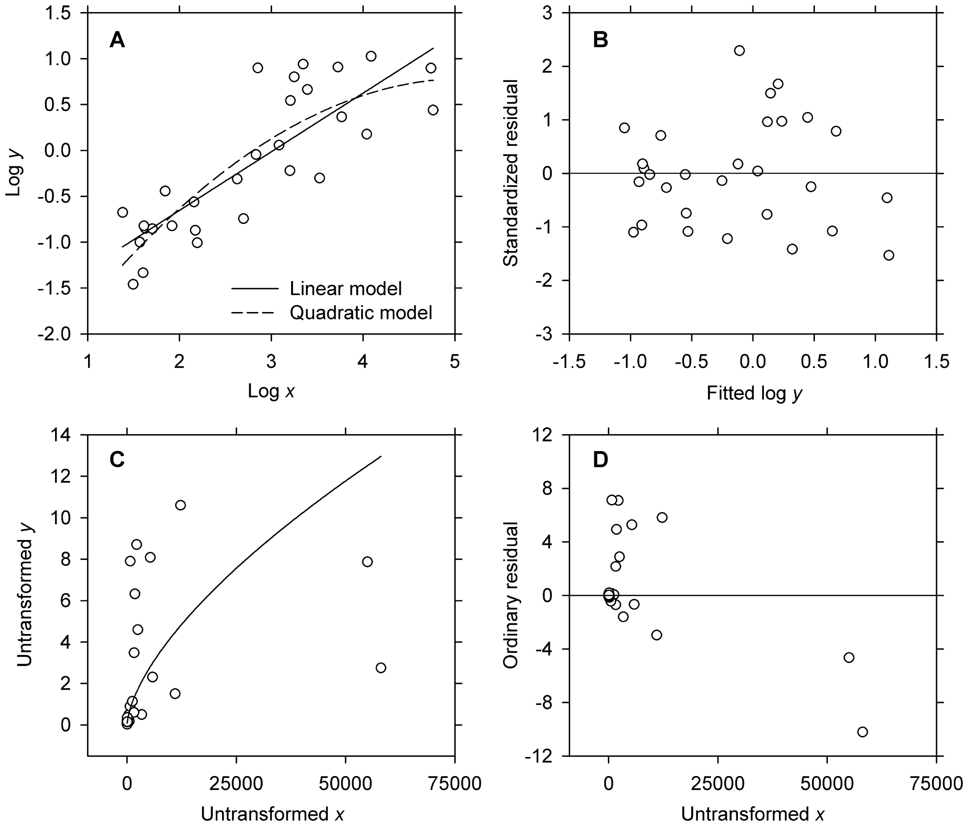

Figure 1. A, linear and quadratic models fitted to logarithmic transformations of bivariate data in dataset #6 (Xiao *et al.*, 2011). The quadratic is not significantly better than the linear model ($F_{1,26} = 3.05$, $P = 0.09$, by nested analysis of variance). B, standardized residuals from the linear model fitted to logarithms are randomly distributed in relation to fitted values and none is so extreme as to mark it as an outlier. C, observations expressed on the arithmetic scale lack a unitary pattern. The mean function estimated by the traditional allometric method does not describe the distribution adequately because it consists of two distinct groupings. D, a plot of ordinary residuals from the traditional allometric model versus untransformed x reinforces the perception that the data were unsuited from the outset for use in an allometric analysis.

back-transforming the equation fitted to logarithms does not describe the arithmetic distribution and, consequently, has no value for prediction or as the basis for theoretical constructs (Fig. 1C). This contention is supported by the decidedly unbalanced distribution for ordinary residuals from the traditional fit (Fig. 1D). Thus, the low AIC for the model fitted by traditional allometry was deceptive: it simply identified the best of a generally bad set of models for describing the original bivariate distribution. The apparent superiority of the traditional fit over the two-parameter model estimated by standard nonlinear regression is a finding of little importance when neither of the models identifies a dominant pattern in the observations.

EXAMPLE 2 (DATASET #386)

The distribution for logarithmic transformations of the 91 observations in dataset #386 satisfied the requirement for linearity (Fig. 2A) but failed to meet the assumption of homoscedasticity (Fig. 2B). Although confidence limits for parameters in the linear model consequently are suspect, the mean function itself probably is reliable (as demonstrated by balance in the distribution of positive and negative residuals). Aside from the question about heteroscedasticity, the logarithmic distribution appeared to be suitable for traditional allometric analysis.

AIC indicated that the power function estimated by the traditional method is substantially better than the two-parameter power function with additive, normal, homoscedastic error (Table 2). This finding is consistent with the conclusion by Xiao *et al.* (2011) that the traditional method generally is better than standard nonlinear regression for fitting allometric equations. However, a two-parameter model with normal, heteroscedastic error actually has the lowest AIC and apparently is the best model in the pool of candidates (Table 2).

Table 1. Predictive equations and AICs for models fitted to dataset #6 from Xiao *et al*. (2011)

Predictive equation	AIC	ΔAIC
Back-transformed OLS with lognormal, heteroscedastic error: $\hat{y} = 0.0117x^{0.639}$	65.6	0
Linear model (no intercept) with normal, homoscedastic error: $\hat{y} = (1.261e-04)x$	159.5	93.9
Linear model (intercept) with normal, homoscedastic error: $\hat{y} = 1.960 + (7.831e-05)x$	152.6	87.0
Two-parameter power with normal, homoscedastic error: $\hat{y} = 0.405x^{0.261}$	144.2	78.6
Three-parameter power with normal, homoscedastic error: Failed to converge	–	–
Linear model (no intercept) with normal, heteroscedastic error: $\hat{y} = 0.0018x$	109.8	44.2
Linear model (intercept) with normal, heteroscedastic error: $\hat{y} = 0.063 + 0.002x$	113.0	47.4
Two-parameter power with normal, heteroscedastic error: $\hat{Y} = 0.008x^{0.780}$	106.3	40.7
Three-parameter power with normal, heteroscedastic error: $\hat{y} = -0.023 + 0.011x^{0.736}$	107.2	41.6

AIC, Akaike's information criterion; OLS, ordinary least squares.

Table 2. Predictive equations and AICs for models fitted to dataset #386 from Xiao *et al*. (2011)

Predictive equation	AIC	ΔAIC
Back-transformed OLS with lognormal, heteroscedastic error: $\hat{y} = 2.727x^{0.061}$	119.9	5.8
Linear model (no intercept) with normal, homoscedastic error: $\hat{y} = (9.904e-05)x$	504.9	390.8
Linear model (intercept) with normal, homoscedastic error: $\hat{y} = 3.821 + (2.079e-05)x$	179.4	65.3
Two-parameter power with normal, homoscedastic error: $\hat{y} = 2.761x^{0.060}$	130.8	16.7
Three-parameter power with normal, homoscedastic error: Failed to converge	–	–
Linear model (no intercept) with normal, heteroscedastic error: $\hat{y} = (1.122e-04)x$	505.8	391.7
Linear model (intercept) with normal, heteroscedastic error: $\hat{y} = 3.625 + (2.870e-04)x$	176.3	62.2
Two-parameter power with normal, heteroscedastic error: $\hat{y} = 2.727x^{0.061}$	114.1	0
Three-parameter power with normal, heteroscedastic error: $\hat{y} = -9.014 + 11.942x^{0.013}$	125.5	11.4

AIC, Akaike's information criterion; OLS, ordinary least squares.

When untransformed data were displayed graphically, it again became apparent that logarithmic transformation had caused ugly data to look remarkably good. None of the statistical models in the pool of candidates provides an adequate description for pattern in the observations because no unitary pattern exists (Fig. 2C). Three aberrant observations stand out in the sample (Fig. 2C) and they also are apparent in a graph of ordinary residuals from the traditional model (Fig. 2D). As in the preceding

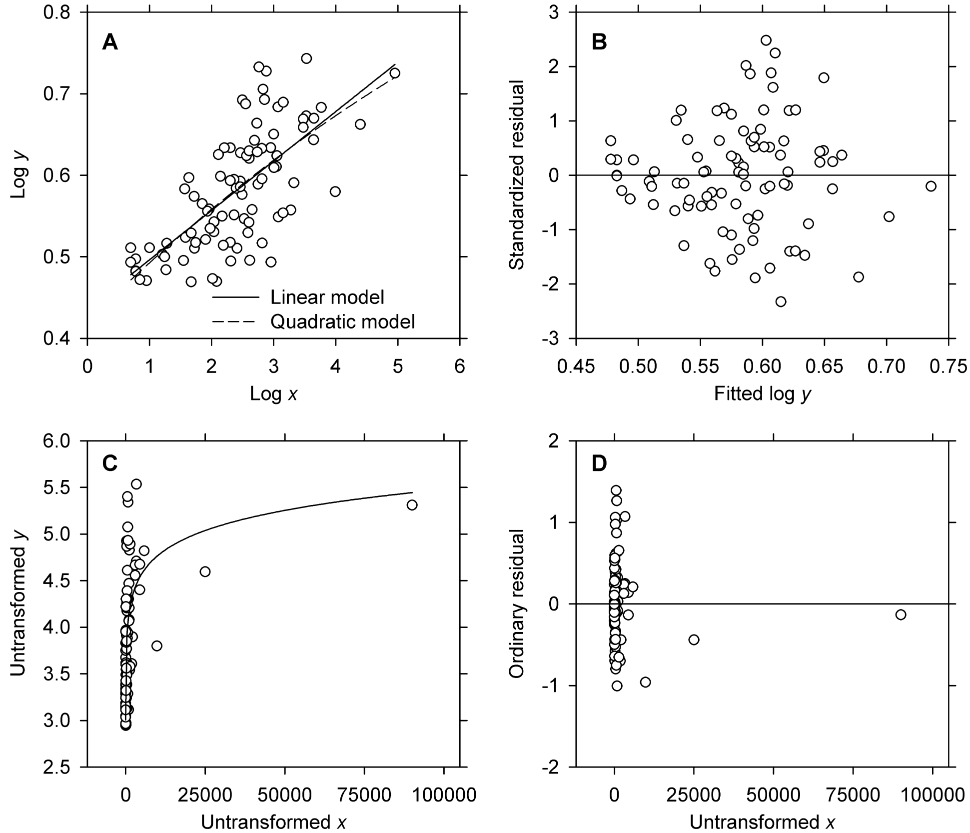

Figure 2. A, linear and quadratic models fitted to logarithmic transformations of bivariate data in dataset #386 (Xiao *et al*., 2011). The quadratic is not significantly better than the linear model ($F_{1,88} = 0.29$, $P = 0.59$, by nested analysis of variance). B, standardized residuals from the linear model fitted to transformations show a megaphone-shaped pattern indicative of heteroscedasticity, but negative and positive residuals are balanced. C, three observations on the original scale stand apart from the distribution for the other 88 observations in the sample. The data are unsuited for allometric analysis and cannot be described by the model fitted by traditional allometry. D, a plot of ordinary residuals from the traditional allometric model versus untransformed *x* reinforces the perception that the data were unsuited from the outset for use in an allometric analysis.

example, the original data are not well suited for use in bivariate allometry. The low quality of the data escaped detection by Xiao *et al*. (2011) because the traditional model apparently was not validated in the scale of measurement (Fig. 2C). The allometric exponent is unreliable, and the mean function estimated by traditional allometry has little value for prediction. The procedure followed by Xiao *et al*. (2011) failed to detect the flaw in the dataset and led to the trivial and misleading conclusion that a traditional model is better than a two-parameter model fitted by standard nonlinear regression.

Example 3 (dataset #312)

Logarithmic transformation of the 27 observations in dataset #312 failed to linearize the distribution (Fig. 3A), but the cause for the failure appears to have been a single, outlying observation at the low end of the distribution (Fig. 3B). Inasmuch as the departure from linearity in log domain violates a requisite condition for application of the traditional allometric method (Reeve, 1940; Kavanagh & Richards, 1942; Richards & Kavanagh, 1945), this unusual observation probably should have been reconciled before studying the full dataset (Osborne & Overbay, 2004). Nevertheless, the distribution apparently was judged by Xiao *et al*. (2011) to be approximately linear because they appear to have used all 27 observations in their analysis. Examination of the data therefore continued.

AIC for the two-parameter power function fitted by standard nonlinear regression was substantially lower than AIC for the traditional model (Table 3). This dataset apparently was one of the few studied by Xiao *et al*. (2011) where nonlinear regression appeared to perform better than the traditional allometric method. However, such a conclusion is misleading because the

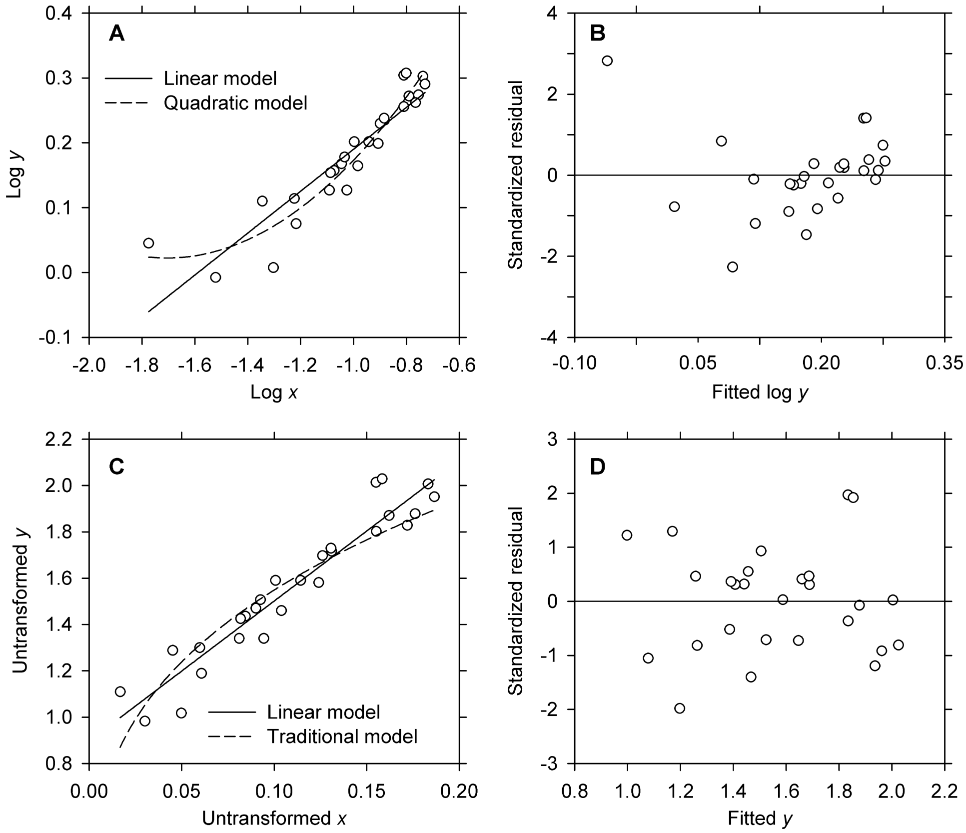

Figure 3. A, linear and quadratic models fitted to logarithmic transformations of bivariate data in dataset #312 (Xiao *et al*., 2011). The quadratic is better than the linear model ($F_{1,24}$ = 22.81, $P < 0.001$, by nested analysis of variance; adjusted R^2 = 0.83 and 0.91 for linear and quadratic models, respectively). The curvilinearity may be caused by an outlier at the low end of the distribution. B, standardized residuals from the linear model fitted to transformations point to the presence of an outlier at the low end of the distribution for fitted values. The Studentized deleted residual for this unusual observation (5.05) supports the contention that it was not drawn from the same bivariate distribution as the other 26 observations (Kutner *et al*., 2004). C, linear model with additive, normal, homoscedastic error provides the best description for pattern in the untransformed observations. The traditional allometric model overestimates observations in the middle of the distribution and underestimates them at the extremes. D, standardized residuals for the linear model fitted to untransformed data are randomly distributed with respect to fitted values for y.

best model in the candidate pool actually is a straight line with a nonzero intercept and additive, normal, homoscedastic error (Table 3). The estimate for the allometric exponent in the best model consequently is 1. The mean function for the straight line captures the dominant pattern in the observations (Fig. 3C) and residuals are randomly distributed with respect to fitted values (Fig. 3D). A normal probability plot approximates a straight line with intercept of 0 and slope of 1 (see Supporting information, Fig. S1), so residuals conform to expectation for a normal distribution (Ritz & Streibig, 2008). The distribution of residuals is also noteworthy because no outlier is evident (Fig. 3D). Thus, the appearance of an outlier in the log-log distribution apparently was an artefact of transformation (Fig. 3B).

Two other models are statistically equivalent to the best model in the candidate pool (Table 3). The three-parameter power function with additive, normal, homoscedastic error approximates a straight line with parameter estimates that are almost identical to those for the best model. The straight line with additive, normal, heteroscedastic error also has regression coefficients similar to those of the best fit, but the coefficient θ in this second equivalent model (−0.24) points to variance that may decline slightly with increasing $\hat{y}$. In any event, all evidence points to a rectilinear model (with intercept) as the best in the candidate pool, and all three of the favoured models have additive, normal error.

The mean function estimated by the traditional allometric method does not follow the path of the

Table 3. Predictive equations and AICs for models fitted to dataset #312 from Xiao *et al*. (2011)

Predictive equation	AIC	ΔAIC
Back-transformed OLS with lognormal, heteroscedastic error: $\hat{y} = 3.258x^{0.323}$	–28.9	20.0
Linear model (no intercept) with normal, homoscedastic error: $\hat{y} = 12.918x$	26.5	75.4
Linear model (intercept) with normal, homoscedastic error: $\hat{y} = 0.897 + 6.047x$	–48.9	0
Two-parameter power with normal, homoscedastic error: $\hat{y} = 3.611x^{0.368}$	–38.3	10.6
Three-parameter power with normal, homoscedastic error: $\hat{y} = 0.896 + 6.041x^{0.999}$	–46.9	2.0
Linear model (no intercept) with normal, heteroscedastic error: $\hat{y} = 12.015x$	18.0	66.9
Linear model (intercept) with normal, heteroscedastic error: $\hat{y} = 0.898 + 6.039x$	–47.0	1.9
Two-parameter power with normal, heteroscedastic error: $\hat{y} = 3.967x^{0.413}$	–43.2	5.7
Three-parameter power with normal, heteroscedastic error: $\hat{y} = 0.875 + 5.740x^{0.960}$	–45.0	3.9

AIC, Akaike's information criterion; OLS, ordinary least squares.

observations in arithmetic domain and, consequently, is an unreliable basis for prediction and theoretical constructs (Fig. 3C). The traditional model is inadequate because it lacks an intercept and because it has an inappropriate form for error (i.e. lognormal instead of normal). The two-parameter model fitted by standard nonlinear regression also suffers from underparameterization. These problems were not detected with the protocol recommended by Xiao *et al*. (2011).

Example 4 (dataset #471)

The bivariate distribution for the 112 observations comprising dataset #471 deviated somewhat from linearity in log space, largely as a result of the influence of a small cluster of points at the high end of the distribution (Fig. 4A). However, this departure from linearity is not obvious in the plot of standardized residuals (Fig. 4B) and fitting a quadratic model resulted in only a marginal improvement in fit ($R^2 = 0.94$ and 0.95 for the linear and quadratic models, respectively). The appearance of curvilinearity may have been nothing more than a statistical anomaly, so the premise that the distribution was effectively linearized by transformation was accepted. The alternative at this point would have been to invoke non-loglinearity, which is a concept of questionable utility (Packard, 2012). The model for the straight line therefore was back-transformed to form a two-parameter power model in the arithmetic scale (Table 4).

AIC for the traditional model was substantially lower than that for the two-parameter model with additive, normal, homoscedastic error (Table 4), so this case presumably was another that led Xiao *et al*. (2011) to their conclusion about the overall superiority of the traditional allometric method. Nevertheless, the best of the candidate models is the three-parameter function with additive, normal, heteroscedastic error (Table 4). The mean function follows the path of the observations much better than does the function estimated by back-transformation (Fig. 4C) and normalized residuals for the three-parameter power model are acceptable (Fig. 4D). A normal probability plot approximates a straight line with slope of 1 and intercept of 0 (see Supporting information, Fig. S2), so the analysis satisfies the assumption of normality in the distribution for residuals (Ritz & Streibig, 2008). The best estimate for the allometric exponent consequently is 0.80 and not the 0.72 that is predicted by the traditional method. The major cause for this difference in exponents is the inclusion of an explicit intercept in the regression model. Allometric equations commonly need an intercept to capture the dominant pattern in the bivariate distribution (Verwijst, 1991; Bales, 1996; Sartori & Ball, 2009; Packard, 2012). Underparameterization was not detected by the procedure articulated by Xiao *et al*. (2011).

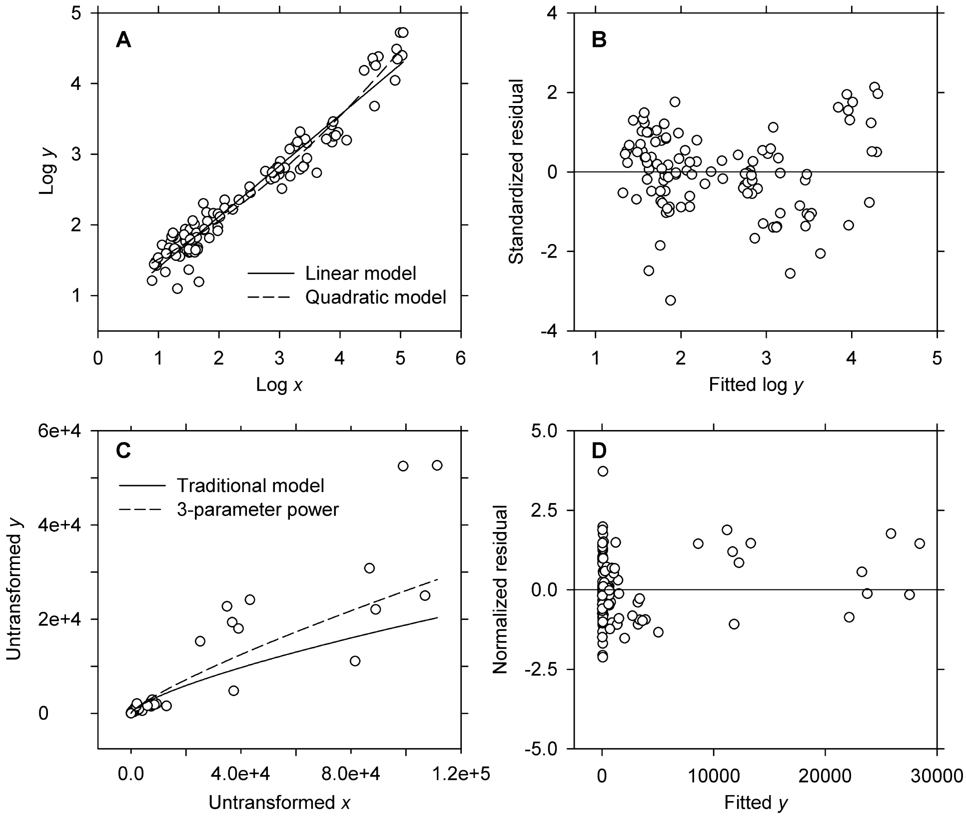

Figure 4. A, linear and quadratic models fitted to logarithmic transformations of bivariate data in dataset #471 (Xiao *et al.*, 2011). The quadratic is better than the linear model ($F_{1,109} = 18.80$, $P < 0.001$, by nested analysis of variance). B, standardized residuals from the linear model fitted to transformations are not obviously flawed. C, a three-parameter power model with additive, normal, heteroscedastic error provides the best description for the pattern in untransformed observations. The traditional allometric model underestimates *y* in the middle and upper parts of the distribution. D, normalized residuals for the three-parameter power model point to a possible outlier at the low end of the distribution for fitted values but the remainder of the distribution is satisfactory.

DISCUSSION

The recent controversy over how to fit a two-parameter power function to bivariate data has raised new concerns about the reliability of prior research that was based on statistical models fitted by the traditional allometric method. Xiao *et al.* (2011) set out to resolve the debate by re-examining 471 datasets taken from the literature. A model fitted by the traditional method was judged by Xiao *et al.* (2011) to be better than one fitted by standard nonlinear regression for describing 69% of the datasets, and the two approaches were said to be equally good in another 15% of the cases. Thus, the traditional fit was supported in a large majority of the comparisons. However, the few cases in which standard nonlinear regression appeared to yield the better model (17%) led Xiao *et al.* (2011) also to recommend that investigators consider the probable form for error before settling on a modelling procedure. If empirical analysis (i.e. AIC) or theoretical considerations point to 'lognormal' error for the model, the traditional allometric method was recommended. Otherwise, standard nonlinear regression was said to be the procedure of choice.

This advice from Xiao *et al.* (2011) is of questionable value because important steps were omitted from their analyses and because their protocol was focused too narrowly on a pair of two-parameter models (also Mascaro *et al.*, 2011, 2014; Ballantyne, 2013; Lai *et al.*, 2013). For example, Xiao *et al.* (2011) did not perform exploratory analyses on untransformed data in bivariate display (see Packard, 2013), else they most certainly would have detected at the outset flawed datasets such as those in Examples 1 and 2 (Figs 1C, 2C). Moreover, by limiting their pool of candidate models to a pair of two-parameter power functions with different forms for error, Xiao *et al.* (2011) sometimes fitted underparameterized functions that failed to capture the dominant pattern in

Table 4. Predictive equations and AICs for models fitted to dataset #471 from Xiao *et al*. (2011)

Predictive equation	AIC	ΔAIC
Back-transformed OLS with lognormal, heteroscedastic error: $\hat{y} = 4.775x^{0.719}$	1412.3	12.8
Linear model (no intercept) with normal, homoscedastic error: $\hat{y} = 0.361x$	2153.7	754.2
Linear model (intercept) with normal, homoscedastic error: $\hat{y} = 143.7 + 0.359x$	2155.6	756.1
Two-parameter power with normal, homoscedastic error: $\hat{y} = 0.858x^{0.924}$	2155.0	755.5
Three-parameter power with normal, homoscedastic error: $\hat{y} = 10.278 + 0.848x^{0.925}$	2157.0	757.5
Linear model (no intercept) with normal, heteroscedastic error: $\hat{y} = 0.420x$	1526.2	126.7
Linear model (intercept) with normal, heteroscedastic error: $\hat{y} = 46.767 + 0.477x$	1425.4	25.9
Two-parameter power with normal, heteroscedastic error: $\hat{y} = 5.736x^{0.701}$	1414.4	14.9
Three-parameter power with normal, heteroscedastic error: $\hat{y} = 21.116 + 2.487x^{0.804}$	1399.5	0

AIC, Akaike's information criterion; OLS, ordinary least squares.

untransformed data (Figs 3C, 4C). Finally, Xiao *et al.* (2011) apparently did not routinely validate fitted models graphically on the arithmetic scale. This was a critically important oversight. The primary objective in fitting any allometric model is to estimate a mean function that captures the dominant pattern in the untransformed data, yet success in achieving this goal can be judged only by examining a graphical display of the mean function against the background of untransformed observations (Anscombe, 1973; Kutner, Nachtsheim & Neter, 2004; Cook & Weisberg, 2009). Examining AIC is not sufficient to identify a good statistical model because all the models in the candidate pool may be bad (Burnham & Anderson, 2002: 62). Reliance on AIC is especially inappropriate when one model is fitted directly to untransformed observations and the other is fitted indirectly by back-transforming from the logarithmic scale (Packard, 2013).

The narrow focus for the investigation also led Xiao *et al*. (2011) to conclude prematurely that lognormal error is more common in nature than normal error (see also Mascaro *et al*., 2011, 2014; Ballantyne, 2013; Lai *et al*., 2013). Both the datasets that were amenable to statistical analysis in the current investigation were described best by models with normal error (see Supporting information, Figs S1, S2), despite the fact that one of those datasets (#471) presumably was characterized by Xiao *et al*. (2011) as having a lognormal distribution for residuals (Table 4). The characterization of dataset #471 as having lognormal error most likely was caused by confounding heteroscedasticity with lognormality: the data were heteroscedastic but not lognormal. Such confounding of heteroscedasticity with lognormality is not uncommon (Packard, 2013, 2014b).

The present study examined 82 datasets chosen haphazardly from the 471 compiled by Xiao *et al.* (2011), but the results are difficult to summarize because of the relatively large number of models in the candidate pool and because more than one model was an acceptable fit to some datasets (Table 3). Some datasets are described quite well by a two-parameter power function fitted by the traditional allometric method (multiplicative, lognormal, heteroscedastic error); others are well characterized by a two-parameter power function fitted by standard nonlinear regression (additive, normal, homoscedastic error); and many are not described adequately by either of these procedures (e.g. as in the four case studies in the current investigation). The point here is that recommendations tendered by Xiao *et al*. (2011) do not provide workers with a reliable guide for performing research on allometric variation. A more holistic approach to allometric analysis is needed in future investigations to avoid problems such as those described in the present study. Investigators should identify larger pools of candidate models, with different functional form and different structure for error. The traditional allometric method should be included in the mix because the resultant model might be quite appropriate when data are heteroscedastic and residuals are demonstrably lognormal. However, the resulting mean function then will predict geometric means for the response instead of arithmetic means (Smith, 1993; Hayes & Shonkwiler, 2006). Regardless of the method that is used to fit the model, it should be validated graphically in the scale of measurement. Neither AIC nor any other procedure for identifying the 'best' model can substitute for graphical display.

ACKNOWLEDGEMENTS

I thank Geoff Birchard for sharing the results of his preliminary study of transformed observations in the 471 datasets compiled by Xiao *et al*. (2011). I am also grateful to Ethan White and four anonymous referees for suggestions and criticisms that enabled me to improve the account appreciably. Special thanks, however, go to the editor, Professor John Allen, who was willing to work with me through several revisions of a manuscript that was unpopular with some of the reviewers.

REFERENCES

Anscombe FJ. 1973. Graphs in statistical analysis. *American Statistician* **27:** 17–21.

Bales GS. 1996. Heterochrony in brontothere horn evolution: allometric interpretations and the effect of life history scaling. *Paleobiology* **22:** 481–495.

Ballantyne F. 2013. Evaluating model fit to determine if logarithmic transformations are necessary in allometry: a comment on the exchange between Packard (2009) and Kerkhoff and Enquist (2009). *Journal of Theoretical Biology* **317:** 418–421.

Burnham KP, Anderson DR. 2002. *Model selection and multimodel inference*, 2nd edn. New York, NY: Springer.

Cawley GC, Janacek GJ. 2010. Letter: on allometric equations for predicting body mass of dinosaurs. *Journal of Zoology* **282:** 223–225.

Cook RD, Weisberg S. 2009. *Applied regression including computing and graphics*. New York, NY: Wiley.

Hayes JP, Shonkwiler JS. 2006. Allometry, antilog transformations, and the perils of prediction on the original scale. *Physiological and Biochemical Zoology* **79:** 665–674.

Huxley JS. 1932. *Problems of relative growth*. London: Methuen, Available at: http://www.archive.org/details/problemsofrelati033234mbp

Kavanagh AJ, Richards OW. 1942. Mathematical analysis of the relative growth of organsisms. *Proceedings of the Rochester Academy of Science* **8:** 150–174.

Kerkhoff AJ, Enquist BJ. 2009. Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology* **257:** 519–521.

Kutner MH, Nachtsheim CJ, Neter J. 2004. *Applied linear regression models*, 4th edn. Boston, MA: McGraw-Hill Irwin.

Lai J, Yang B, Lin D, Kerkhoff AJ, Ma K. 2013. The allometry of coarse root biomass: log-transformed linear regression or nonlinear regression? *PLoS ONE* **8:** e77007.

Marshall DJ, Bode M, White CR. 2013. Estimating physiological tolerances – a comparison of traditional approaches to nonlinear regression techniques. *Journal of Experimental Biology* **216:** 2176–2182.

Mascaro J, Litton CM, Hughes RF, Uowolo A, Schnitzer SA. 2011. Minimizing bias in biomass allometry: model selection and log-transformation of data. *Biotropica* **43:** 649–653.

Mascaro J, Litton CM, Hughes RF, Uowolo A, Schnitzer SA. 2014. Is logarithmic transformation necessary in allometry? Ten, one-hundred, one-thousand-times yes. *Biological Journal of the Linnean Society* **111:** 230–233.

Osborne JW, Overbay A. 2004. The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessessment, Research & Evaluation* **9:** 1–12.

Packard GC. 2009. On the use of logarithmic transformations in allometric research. *Journal of Theoretical Biology* **257:** 515–518.

Packard GC. 2011. Unanticipated consequences of logarithmic transformation in bivariate allometry. *Journal of Comparative Physiology B* **181:** 841–849.

Packard GC. 2012. Is non-loglinear allometry a statistical artifact? *Biological Journal of the Linnean Society* **107:** 764–773.

Packard GC. 2013. Is logarithmic transformation necessary in allometry? *Biological Journal of the Linnean Society* **109:** 476–486.

Packard GC. 2014a. Multiplicative by nature: logarithmic transformation in allometry. *Journal of Experimental Zoology B* **322:** 202–207.

Packard GC. 2014b. Assessing allometric growth by leaves and the hypothesis of diminishing returns. *International Journal of Plant Sciences* **175:** 742–753.

Packard GC, Boardman TJ, Birchard GF. 2010. Allometric equations for predicting body mass of dinosaurs: a comment on Cawley & Janacek (2010). *Journal of Zoology* **282:** 221–222.

Pinheiro JC, Bates DM. 2000. *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.

Reeve ECR. 1940. Relative growth in the snout of anteaters. A study in the application of quantitative methods to systematics. *Proceedings of the Zoological Society of London A* **110:** 47–80.

Richards OW, Kavanagh AJ. 1945. The analysis of growing form. In: Le Gros Clark WE, Medawar PB, eds. *Essays on growth and form presented to D'Arcy Wentworth Thompson*. Oxford: Oxford University Press, 188–230.

Ritz C, Streibig JC. 2008. *Nonlinear regression with R*. New York, NY: Springer.

Sartori AF, Ball AD. 2009. Morphology and postlarval development of the ligament of *Thracia phaseolina* (Bivalvia: Thraciidae), with a discussion of model choice in allometric studies. *Journal of Molluscan Studies* **75:** 295–304.

Smith RJ. 1993. Logarithmic transformation bias in allometry. *American Journal of Physical Anthropology* **90:** 215–228.

Verwijst T. 1991. Logarithmic transformations in biomass estimation procedures: violation of the linearity assumption in regression analysis. *Biomass Bioenergy* **1:** 175–180.

Xiao X, White EP, Hooten MB, Durham SL. 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology* **92:** 1887–1894.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. 2009. *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Normal probability plot from the fit of a straight line with additive, normal, homoscedastic error to observations in dataset #312 from Xiao *et al*. (2011). The solid line has a slope of 1 and an intercept of 0. The scatter of points around the line points to a normal distribution for residuals.
Figure S2. Normal probability plot from the fit of a three-parameter power function with additive, normal, heteroscedastic error to observations in dataset #471 from Xiao *et al*. (2011). The solid line has a slope of 1 and an intercept of 0. The scatter of points around the line points to a normal distribution for residuals.
Table S1. Computer code.