2011

On the Use of Log-Transformation vs. Nonlinear Regression for Analyzing Biological Power-Laws

Xiao Xiao
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports
Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
https://digitalcommons.usu.edu/gradreports/1219

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
On the Use of Log-Transformation vs. Nonlinear Regression for Analyzing Biological Power-Laws

By

Xiao Xiao

A report submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE in

Statistics

UTAH STATE UNIVERSITY
Logan, Utah

2011
ABSTRACT

On the Use of Log-Transformation vs. Nonlinear Regression for Analyzing Biological Power-Laws

by

Xiao Xiao, Master of Science
Utah State University, 2011

Major Professor: Dr. Mevin Hooten
Department: Mathematics & Statistics

Power-law relationships are among the most well-studied functional relationships in biology. Recently the common practice of fitting power-laws using linear regression on log-transformed data (LR) has been criticized, calling into question the conclusions of hundreds of studies. It has been suggested that nonlinear regression (NLR) is preferable, but no rigorous comparison of these two methods has been conducted. Using Monte Carlo simulations we demonstrate that the error distribution determines which method performs better, with LR better characterizing data with multiplicative lognormal error and NLR better characterizing data with additive normal error. Analysis of 471 biological power-laws shows that both forms of error occur in nature. While previous analyses based on log-transformation appear to be generally valid, future analyses should choose methods based on the error distribution. We provide detailed guidelines and associated computer code for doing so, including a model averaging approach for cases where the error structure is uncertain.

(52 pages)
ACKNOWLEDGMENTS

I would like to thank my coauthors, Dr. Ethan White, Dr. Mevin Hooten, and Dr. Susan Durham, for the effort they put in for this project. I would also like to thank my committee member, Dr. James Powell, for his support and insightful comments. I am grateful to all the researchers whose papers have been used in data compilation in this study for making their data publicly available; specifically, Dr. Kristina J. Anderson, Dr. Ken Ashwell and Dr. Robin Warne have provided additional help with their data. I would like to thank Dr. Philip Gingerich and Dr. James Haefner for the helpful discussions.

This research is supported financially by the USU College of Science Diversity Fellowship.

Xiao Xiao
# CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>ANALYSIS, RESULTS AND DISCUSSION</td>
<td>4</td>
</tr>
<tr>
<td>- Error Distribution Determines the Best Method for Fitting Power-Laws</td>
<td>4</td>
</tr>
<tr>
<td>- Error Forms Observed in Nature</td>
<td>5</td>
</tr>
<tr>
<td>- Model Averaging: An Alternative Approach When Error Form Is Uncertain</td>
<td>7</td>
</tr>
<tr>
<td>- General Guidelines for the Analysis of Biological Power-Laws</td>
<td>8</td>
</tr>
<tr>
<td>- Implications for Previous Studies</td>
<td>9</td>
</tr>
<tr>
<td>- Complexities</td>
<td>11</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>14</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>15</td>
</tr>
<tr>
<td>FIGURES</td>
<td>19</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>22</td>
</tr>
</tbody>
</table>

| Appendix A | Additional Analysis: Simulation With Hypothetical Parameters | 23 |
| Appendix B | Empirical Datasets | 27 |
| Appendix C | Detailed Simulation Procedures | 39 |
| Appendix D | R Code to Carry Out Guidelines | 41 |
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>List of papers from which empirical allometric datasets were compiled</td>
<td>37</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure | Page
-------|------
1      | An illustration of additive normal error and multiplicative log-normal error displayed on both scales 19
2      | Comparison of bias in point estimation and CI coverage among LR, NLR and AICc-weighted average models 20
3      | Examples from previous analyses of biological power-laws 21
A1     | Effect of sample size, exponent $b$, standard deviation $\sigma$ and error structure on model performance 25
INTRODUCTION

Power-law relationships of the form $y = ax^b$ are one of the most common patterns in biology. They have been documented in a variety of different areas including the relationships between body size, physiological rates and life history traits (Brown et al. 2004), the scaling between body parts in morphology (Farlow et al. 1995), and the species-area relationship in biogeography (Martin & Goldenfeld 2006). These fitted relationships have been used to test the validity of biological theories (Brown et al. 2004), to infer the characteristics of extinct species (Farlow et al. 1995), to assess the effect of evolutionary processes (Mortola & Limoges 2006), and to predict the consequence of habitat loss on biodiversity (Brooks et al. 2002).

Conventional analysis of power-law data uses the fact that log-transforming both sides of the equation yields a linear relationship, $\log(y) = \log(a) + b\log(x)$, allowing log-transformed data to be modeled using linear regression. However, it has been suggested that analysis on logarithmic scales is flawed and that instead, analysis should be carried out on the original scale of measurement using nonlinear regression (Fattorini 2007; Packard & Birchard 2008; Packard 2009; Caruso et al. 2010; Packard et al. 2010). If these claims are correct, the validity of decades of published research in ecology, evolution and physiology would be called into question.

One of the fundamental differences between linear regression on log-transformed data (hereafter, LR) and nonlinear regression on untransformed data (hereafter, NLR) lies in the assumptions about how stochasticity manifests in the model (Gingerich 2000,
Kerkhoff & Enquist 2009). In NLR, it is assumed that the error term is normally
distributed and additive on arithmetic scale (Bates & Watts 1988; Ritz & Streibig 2008):
\[ y = ax^b + \varepsilon, \varepsilon \sim N(0, \sigma^2) \]  
(1)
In contrast, LR assumes that error is normally distributed and additive on the logarithmic
scale (Kerkhoff & Enquist 2009):
\[ \log(y) = \log(a) + b\log(x) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \]
which corresponds to log-normally distributed, multiplicative error on the arithmetic
scale:
\[ y = ax^b e^\varepsilon, \varepsilon \sim N(0, \sigma^2) \]  
(2)
For a single dataset, both assumptions cannot be correct. Violation of statistical
assumptions of error can lead to biased point estimates as well as inaccurate confidence
intervals (Bates & Watts 1988; see Figure 1 for an illustration).

Despite its importance in statistical model fitting, the issue of error distribution
has been largely ignored in discussions about best practices for fitting power-laws to data.
While both additive and multiplicative errors have been posited to occur in biological
systems (Kerkhoff & Enquist 2009; Packard 2009; Cawley & Janacek 2010), to our
knowledge there has been no systematic analysis that evaluates how NLR and LR
estimation methods perform on different error structures, or what form the error structure
actually takes in biological systems. This is surprising given the potential implications of
these methodological issues for understanding biological systems and the strong
arguments regarding appropriate methods being made in the literature (e.g., Packard
2009). Here we use Monte Carlo simulations to test the role of error structure on the
performance of the two methods across empirical parameter space. For cases where the better model cannot be clearly determined, we develop an alternative estimation method based on model averaging. Based on these results, detailed guidelines for the analysis of biological power-laws, and computer code for their implementation, are provided.
ANALYSES, RESULTS AND DISCUSSION

Error Distribution Determines the Best Method for Fitting Power-Laws

Previous arguments regarding the performance of different methods have typically been based on empirical data (but see Hui et al. 2010), despite the fact that it is impossible to know the true data generating mechanism. As such, previous studies provide little insight into the best methodological approach. Monte Carlo simulation, where data are simulated from known distributions, allows for a direct comparison between statistically estimated parameters and their true values. Here we implement the Monte Carlo approach based on parameterizations from empirical datasets so that our results will be valid for the range of empirically observed parameter values. Results from these empirically motivated simulations were consistent with standard Monte Carlo simulations based on hypothetical parameterizations (see Appendix A).

We compiled 471 datasets published between 2004 and 2008 in the fields of ecology and evolution where significant power-law relationships were reported. The selected datasets were either morphological or physiological allometries between organismal traits (for details of data selection and the full list of datasets, see Appendix B). To generate the parameters for simulations each empirical dataset was assumed to have: 1) a multiplicative log-normal error structure, and $a$, $b$, and $\sigma$ were estimated with LR (with $\sigma$ estimated as the standard deviation of the residuals); and 2) an additive normal error structure, and the parameters were estimated with NLR. For each dataset, $10^5$ independent simulations were carried out using the estimated parameters under the
assumption of each error structure. Each simulated dataset was analyzed with both LR and NLR, and the performance of the two methods was compared to determine which method had the better point estimation of $a$ and $b$, as well as more accurate confidence interval (CI) coverage measured by the percentage of simulations where the true parameter value falls within the estimated 95% CI. Only 239 datasets generated valid simulations under the assumption of additive error (see Appendix C for technical details on the procedure of the simulations). All simulations and analyses were carried out using R version 2.9.1 (R Development Core Team 2009). The “nlrwr” package (Ritz & Streibig 2008) was used to compute asymptotic CIs for NLR.

Our simulations confirm the importance of correctly identifying the error distribution when fitting statistical models. Among 471 empirical datasets LR outperformed NLR under the assumption of multiplicative error in all of the datasets (100%) for $a$ and 427 datasets (90.7%) for $b$. Similarly, NLR outperformed LR under the assumption of additive error in 196 datasets (82.0%) for $a$, and 238 datasets (99.6%) for $b$ (out of n = 239 valid parameterizations). The method with the appropriate error assumption also had excellent confidence interval (CI) coverage, whereas CI coverage for the inappropriate method was highly variable, reaching levels as low as 0.2 (Figure 2).

**Error Forms Observed in Nature**

Given the critical nature of the error distribution in determining the appropriate method for analyzing power-law data, it is necessary to understand the form of the error distribution in nature. Previous papers have argued for both normal error (Packard 2009) and log-normal error (Kerkhoff & Enquist 2009), but no systematic analysis of biological power-laws has been conducted.
Taking a likelihood approach to compare the appropriateness of the two error forms for the 471 empirical datasets described above, we used Akaike’s information criterion (AIC), which measures the goodness of fit of a statistical model by incorporating both the likelihood of the model and a penalty for extra parameters (Burnham & Anderson 2002). For each of the 471 empirical datasets, we computed likelihoods and the values of AICc (a second order variant of AIC that corrects for small sample size; see Burnham & Anderson 2002) for both the LR and NLR based models. We compared the AICc values by following the conventional rule that if $|\Delta \text{AICc}|$ (the magnitude of the difference between the two values of AICc) is less than 2, the two models have relatively equal support and cannot be distinguished from each other; otherwise, the model with the lower AICc is considered to have better data support (Burnham & Anderson 2002). Since AICc for the LR model is based on the likelihood from a log-normal distribution conditioned on untransformed data, such comparison does not violate the assumption of identical response variable in AIC-based model selection (Burnham & Anderson 2002, Section 2.11.3).

Consistent with previous suggestions that multiplicative error is biologically more realistic (Gingerich 2000; Kerkhoff & Enquist 2009), our likelihood analysis of 471 allometric datasets shows that log-normal error distributions are substantially more common than normal error distributions, with 68.6% of relationships being better characterized by log-normal error, 16.6% by normal error, and 14.8% having uncertain error structure.
Model Averaging: An Alternative Approach
When Error Form Is Uncertain

Monte Carlo simulations show that if the underlying error structure is known then
the model assuming the appropriate error form (i.e., NLR with normal error, and LR with
log-normal error) will perform well for estimating both the parameters of the power-law
and the CIs of those parameters. However, the underlying error form of real datasets is
not known and our likelihood analysis shows that identification of the error form will not
be clear-cut in all empirical datasets; in part because the error form in real datasets may
be more complex than assumed by the two standard methods. Even in our simulation
models where one distinct error structure has been specified, likelihood tests failed to
identify the correct error structure (i.e., it was either miscategorized or deemed uncertain)
in more than 10% of the simulated datasets for over half of the parameterizations (50.7%
when error was assumed to be log-normal and 71.1% when error was assumed to be
normal).

When two or more models with appreciably different parameter estimates have
similar support, model averaging provides a way to incorporate information from
multiple models so that more stable inference can be made based on the weighted average
of the entire set (Burnham & Anderson 2002; Link & Barker 2006). Several weighting
strategies are available; the most common ones being AIC weight (Burnham & Anderson
2002) and BIC weight (Link & Barker 2006). In our analysis we adopted AIC weight (see
Appendix C for detailed procedure). Based on point estimates and CIs, we assessed
whether the weighted model was able to accurately capture the underlying relationship
under the assumption of the two error structures, i.e. whether it indicated the correct error
structure if one existed. R package “boot” was used to construct CIs for the weighted average model (Davidson & Hinkley 1997; Canty & Ripley 2009).

Comparison of relative bias among LR, NLR and weighted average models shows that the weighted model closely resembles the model with the appropriate error assumption in both point estimation and CI coverage (Figure 2) regardless of error structure. Thus the weighted average model can provide an indication of the appropriate error distribution.

**General Guidelines for the Analysis of Biological Power-Laws**

For future analysis of power-law relationships, we recommend the application of the following two-step procedure to correctly identify and apply the appropriate method:

1. Compare the relative likelihood of the two error structures using AICc or other similar measures. To compute AICc, first fit the two models using NLR and LR respectively and estimate the parameters $a$, $b$, and $\sigma^2$ for each model. Then calculate the likelihood that the data are generated from a normal distribution with additive error

$$L_{norm} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_{NLR}^2}} \exp \left( \frac{-(y_i - a_{NLR}x_{i}^{b_{NLR}})^2}{2\sigma_{NLR}^2} \right),$$

and the likelihood that the data are generated from a log-normal distribution with multiplicative error

$$L_{logn} = \prod_{i=1}^{n} \left( \frac{1}{y_i \sqrt{2\pi \sigma_{LR}^2}} \exp \left( \frac{-(\log(y_i) - \log(a_{LR}x_{i}^{b_{LR}}))^2}{2\sigma_{LR}^2} \right) \right),$$

where $n$ is sample size. AICc for each model can then be computed as $\text{AICc} = 2k -$
\[ 2 \log(L) + \frac{2k(k+1)}{n-k-1}, \]

where \( k \) is the number of parameters (3 in both models) and \( L \) is the corresponding likelihood (Burnham & Anderson 2002).

2a. If \( \text{AIC}_{\text{norm}} - \text{AIC}_{\text{logn}} < -2 \), the assumption of normal error is favored by the data compared to log-normal error, and we can proceed with the results obtained from NLR.

2b. If \( \text{AIC}_{\text{norm}} - \text{AIC}_{\text{logn}} > 2 \), the assumption of log-normal error is favored by the data compared to normal error, and we can proceed with the results obtained from LR.

2c. If \( |\text{AIC}_{\text{norm}} - \text{AIC}_{\text{logn}}| \leq 2 \), neither model is substantially better supported than the other and model averaging should be adopted. The point estimates for \( a \) and \( b \) in the mixed model are then weighted average of the corresponding point estimates from the two original models. The AICc weights of the two models are computed as

\[ w_i = C \cdot \exp \left( - \frac{\text{AICc}_c - \min \left( \text{AICc}_{\text{norm}}, \text{AICc}_{\text{logn}} \right)}{2} \right) \]

where \( C \) is a normalizing constant so that \( w_{\text{norm}} \) and \( w_{\text{logn}} \) sum to 1. CIs for \( a \) and \( b \) can be generated by bootstrapping for dataset of sufficient size (Efron & Tibshirani 1994).

Computer code that implements these recommendations is available in Appendix D.

**Implications for Previous Studies**

For decades LR has been the conventional approach in the analysis of biological power-laws. If the current proposition to replace LR with NLR (e.g., Packard 2009, Packard et al. 2010) were generally legitimate, the conclusions from large numbers of allometric studies would be called into question. However, our likelihood analysis with
471 empirical datasets spanning ecology, evolution and physiology shows that log-normal error consistently provides superior fits to normal error distribution, which implies that the majority of previous allometric studies in these fields using LR are generally valid and contradicts the recent argument that LR is inherently flawed and should be replaced by NLR (e.g., Packard 2009; Packard et al. 2010). As our Monte Carlo simulation studies show, the application of NLR to such datasets will lead to biased parameter estimates and potentially erroneous inferences.

The implications of these results for real biological patterns can be seen by applying the guidelines described in the previous section to arbitrate two debates regarding the exponents of morphological and physiological power-laws. The first example addresses whether or not the scaling of mammalian metabolic rate as a function of body size is consistent with the canonical 0.75 scaling exponent predicted by metabolic theory (Brown et al. 2004). Savage et al. (2004) analyzed one of the most comprehensive datasets on mammalian basal metabolic rate using LR and found that the empirical data supported the predicted form of the relationship ($b_{LR} = 0.74, CI_{0.95} = (0.71, 0.76)$; see Figure 3a). However, reanalyzing the same data using NLR resulted in different parameter estimates and confidence intervals ($b_{NLR} = 0.91, CI_{0.95} = (0.88, 0.94)$), which suggested that the 0.75 exponent be rejected as a reasonable description of the data (Packard & Birchard 2008). A quantitative analysis of the error structure in this dataset shows that the assumption of multiplicative log-normal error is strongly supported compared to additive normal error ($\text{AIC}_{\text{norm}} - \text{AIC}_{\text{logn}} = 306$), suggesting that the data are consistent with the theoretical exponent.

Another example of how this approach can provide clear guidance when LR and
NLR yield different results is the scaling relationship between eye size and brain mass. Burton (2006) analyzed this relationship in fissiped Carnivora using LR and argued that because the exponent did not differ significantly from one ($b_{LR} = 0.87, \text{CI}_{0.95} = (0.55, 1.19)$) that eye size is determined (at least in part) by a simple limitation on the amount of space available in the head. A reanalysis of this data using NLR suggested that bears were outliers and that excluding this taxon the exponent was steeper than the hypothesized value of one ($b_{NLR} = 1.42, \text{CI}_{0.95} = (1.13, 1.70)$). However, both the identification of outliers and the use of nonlinear regression were controversial (Kerkhoff & Enquist 2009). Likelihood analysis demonstrates that the assumption of log-normal error is more strongly supported regardless of whether the bears are included ($\text{AICc}_{\text{norm}} - \text{AICc}_{\text{log}} = 35.9$) or not ($\text{AICc}_{\text{norm}} - \text{AICc}_{\text{log}} = 7.88$). Therefore since LR yields confidence intervals that include one even when the bears are excluded ($b_{LR} = 1.24, \text{CI}_{0.95} = (0.96, 1.53)$; see Figure 3b), the proposed isometric relationship is supported by the data.

Parallel examples where datasets with normal or undetermined error structures suffer from methodological problems are rarer in literature due to the prevalence of log-normal error distribution observed in nature. Nonetheless, reanalysis of the original data is warranted in cases where there is reason to suspect that an additive normal error structure or an undetermined error structure is more realistic.

**Complexities**

Apart from making inferences about the parameters, power functions also are frequently used to make predictions for new observations, which is particularly important in paleontology and conservation biology. For LR, it should be noted that although the parameter estimates are unbiased when the error is log-normal and multiplicative
(Ferguson 1986), the model predicts log(y), and the predicted value of y obtained by anti-log transformation is biased on arithmetic scale (Hayes & Shonkwiler 2006). Measures should be taken to correct for this bias if predictions are to be made from log-transformed power functions (Hayes & Shonkwiler 2006).

One class of commonly observed biological power-law relationships not included in this study is the scaling relationship between species richness and attributes of the habitat (e.g., area, resource availability, distance to mainland, etc.). The most widely studied of these relationships is the species-area relationship (SAR). SARs are of fundamental importance in conservation biology where they are used for making predictions regarding the effect of habitat loss on biodiversity (Brooks et al. 2002) as well as the identification of hotspots (Veech 2000). Studies have shown that inferences on SARs vary with the model that is being fitted (Fattorini 2007). One often overlooked characteristic of SARs is that the response variable, species richness, is a discrete count, which in principle cannot be accommodated by either LR or NLR because both assume a continuous data distribution (which is why this type of data was not included in our empirical analyses). However, the existence of this discrete error structure in biological power-laws suggests that additive normal error and multiplicative log-normal error are often not the only two options that should be considered when analyzing error distributions. O’Hara and Kotze (2010) showed that ignorance of the error characteristics can lead to failure of the statistical analysis. Our understanding of the validity of previous studies of SARs and other relationships that potentially violate the distributional assumptions of LR and NLR would be enhanced by a systematic comparison between methods that accommodate their statistical and biological properties and currently applied
methods such as NLR and LR.
CONCLUSIONS

Power functions are one of the most broadly studied relationships in biological systems. The current debate surrounding the methodology used in their analysis has generated considerable confusion in the field. As a result the conclusions of previous studies have been called into question and the progress of new analyses has been hampered. Our study provides a clear answer to the current controversy surrounding the appropriate methodology for analyzing allometric data. Neither linear regression on log-transformed data nor standard nonlinear regression is inherently superior for fitting power-laws to data. Which method performs better depends on the distribution of the error. For most allometric datasets like those we studied, the error is distributed such that log-transformed linear regression will produce more accurate parameter estimates and confidence intervals; thus most published results likely are valid. However, the methodology chosen for future analyses of power-laws in ecology and evolution should be based on explicit analyses of the underlying error structure. We recommend that likelihood comparisons be applied to assess the error structure of the dataset. In cases where the error is approximately multiplicative lognormal, the log-transformed linear regression should be used, while nonlinear regression on untransformed data should be applied to those datasets with additive normal error. For datasets with an indeterminate error structure, we recommend using model averaging to calculate the weighted average of the parameter estimates.
REFERENCES


Fattorini, S. 2007. To fit or not to fit? A poorly fitting procedure produces inconsistent results when the species-area relationship is used to locate hotspots. Biodiversity and Conservation 16: 2531-2538.


FIGURES

Figure 1. An illustration of additive normal error and multiplicative log-normal error displayed on both arithmetic and logarithmic scales, and how the underlying relationships can be distorted by the application of the inappropriate methods. For additive error, $x$ was generated from a uniform distribution ranging from 10 to 10000, $y$ was generated using Eqn.1 with $a = 10, b = 0.2, \sigma = 10$. For multiplicative error, $x$ was generated from a log-uniform distribution ranging from 1 to 10 on logarithmic scale, $y$ was generated using Eqn.2 with $a = 0.3, b = 0.75, \sigma = 0.3$. The dashed curves correspond to the true underlying relationships.
Figure 2. Comparison of bias in point estimation and CI coverage among LR, NLR and AICc-weighted average models in simulations with parameters estimated from 471 empirical datasets for multiplicative error structure and 239 empirical datasets for additive error structure. Relative bias (mean estimate/true value) is depicted because a spans a wide range across empirical datasets. For point estimation, each curve represents the relative frequency distribution of relative bias. An appropriate method peaks at 0 (on logarithmic scale) with small dispersion, while an inappropriate method shows a wide range of relative bias. For CI coverage, the horizontal dashed line represents the nominal 0.95 level. Note that point estimates were generated based on $10^5$ simulated datasets, while CIs were based on 400 additional simulated datasets due to computational limitation. Only the results for $b$ are shown.
Figure 3. Examples from previous analyses of biological power-laws where misapplication of methods have led to inaccurate characterization of the relationships: a. BMR – body mass relationship taken from Savage et al. 2004, analyzed in Packard & Birchard 2008; b. eye size – brain mass relationship taken from Burton 2006, analyzed in Packard 2009. See text for details.
APPENDICES
Appendix A  Additional Analysis: Simulation
With Hypothetical Parameters

We compared the performance of LR and NLR under a variety of different scenarios comparable to the parameter space of realistic biological relationships. In addition to error structure, we also evaluated the effects of three other variables (exponent $h$, sample size, and standard deviation of the error $\sigma$) which could influence model performance. The exponent took one of nine values from -1 to 1 with interval 0.25, encompassing both positive and negative correlations between $x$ and $y$. Sample size was set at one of five levels (20, 50, 100, 200 and 500) which span the size range of most empirical allometric datasets. The standard deviation also had five values, with different values for the two error structures. For normal additive error, $\sigma$ varied from 1 to 5 with interval 1. For log-normal multiplicative error, $\sigma$ varied from 0.1 to 0.5 with interval 0.1.

For each combination of these three variables $10^5$ simulations were carried out, assuming first a normal error structure and then a log-normal error structure. Coefficient $a$ in the two equations was set to be 5 in the case of normal error and log(5) in the case of log-normal error. In each simulation, $x$ was randomly generated from a log-uniform distribution ranging from 10 to $10^5$, and $y$ was computed using Eq.1 for normal error or Eq.2 for log-normal error, with the stochastic term $\epsilon$ randomly sampled from a normal or a log-normal distribution, respectively. While the assumption of multiplicative log-normal error always generated feasible datasets with positive $y$ values, some parameterizations using additive normal error led to simulated datasets with negative $y$, which we deemed to be biologically unrealistic. We thus discarded any parameterization for which more than 10% of the simulated datasets included negative $y$ values; for the
retained parameterizations, we discarded any simulated dataset with negative y values. Each simulated dataset was analyzed with both LR and NLR, and the performances of the two methods were compared based on point estimation and CI coverage.

Our simulations with hypothetical parameters further validate our conclusion that error distribution is critical for model performance. Figure A1 shows the average absolute deviation of the point estimates from the true parameter values and the associated CI coverage. The model that assumed the appropriate error form (i.e., NLR with normal error, and LR with log-normal error) yielded unbiased estimates of parameters and accurate CI coverage regardless of sample size, exponent value and standard deviation. In contrast, the performance of the model that assumed the inappropriate error form fluctuated widely, yielding significantly biased estimates and questionable CI coverage in most situations.
Figure A1. Effect of sample size, exponent $b$, standard deviation $\sigma$ and error structure on model performance, measured by mean absolute deviation from hypothesized parameters and CI coverage. Results for both $a$ and $b$ display similar trends; results for only $b$ are shown. For normal error distribution, all parameterizations with negative $b$ are removed because all result in unrealistic datasets with negative $y$. For log-normal error distribution, levels of $b$ with the same absolute value and opposite signs show the same pattern. Thus we combine the two levels of $b$ with opposite signs by taking the average of both the absolute deviation and CI coverage for log-normal error, and only present the results for 5 levels of $|b|$. Different levels of $\sigma$ are depicted with different dot sizes; sample sizes are represented as sections along the horizontal axis. Within each sample size section, $|b|$ increases from left to right in 5 levels from 0 to 1. For CI coverage, the horizontal line
represents the nominal 0.95 level. Note that some parameterizations are missing for normal error due to the removal of unrealistic parameterizations. $|b| = 0$ (i.e. the leftmost dots within each sample size section) leads to realistic parameterizations only when $\sigma = 1$ regardless of sample size. In addition, large sample size combined with high $\sigma$ (sample size 500, $\sigma = 4$ or 5; sample size 200, $\sigma = 4$ or 5; sample size 100, $\sigma = 5$; sample size 50, $\sigma = 5$) yields unrealistic parameterizations when $|b| = 0.25$. 
Appendix B  Empirical Datasets

We performed a literature search on ISI Web of Knowledge using key words “allometry” and “Ecology or Evolution”. We examined all accessible datasets published between 2004 and 2008 that reported significant power-law relationships. We dropped datasets that had less than 5 observations or for which our parameter estimates differed from those reported in the original paper (i.e., datasets that might contain typographical errors). Overall, 471 datasets describing both morphological and physiological relationships met these criteria. All papers included in our analysis are listed below; the number of datasets taken from each paper is summarized in Table B1. The compiled datasets are not included due to the size of the table.


Li, Y.-N., D.-M. Yang, S.-C. Sun, and X.-M. Gao. 2008. Effects of twig size on biomass allocation within twigs and on lamina area supporting efficiency in


Marroig, G. 2006. When size makes a difference: allometry, life-history and morphological evolution of capuchins (Cebus) and squirrels (Saimiri) monkeys (Cebinae, Platyrhini). BMC Evolutionary Biology 7: 20.


Table B1. List of papers from which empirical allometric datasets were compiled.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Dataset ID</th>
<th>Number of Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doucette &amp; Geiser</td>
<td>2008</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Nogueira et al.</td>
<td>2008</td>
<td>2 to 3</td>
<td>2</td>
</tr>
<tr>
<td>Shik</td>
<td>2008</td>
<td>4 to 7</td>
<td>4</td>
</tr>
<tr>
<td>Wiff &amp; Roa-Ureta</td>
<td>2008</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Yang et al.</td>
<td>2008</td>
<td>9 to 15</td>
<td>7</td>
</tr>
<tr>
<td>Lisney et al.</td>
<td>2008</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Isler et al.</td>
<td>2008</td>
<td>17 to 18</td>
<td>2</td>
</tr>
<tr>
<td>Li et al.</td>
<td>2008</td>
<td>19 to 22</td>
<td>4</td>
</tr>
<tr>
<td>Gerstner &amp; Gerstein</td>
<td>2008</td>
<td>23 to 27</td>
<td>5</td>
</tr>
<tr>
<td>Habeck &amp; Meehan</td>
<td>2008</td>
<td>28 to 30</td>
<td>3</td>
</tr>
<tr>
<td>Warne &amp; Charnov</td>
<td>2008</td>
<td>31 to 32</td>
<td>2</td>
</tr>
<tr>
<td>Cooper &amp; Withers</td>
<td>2008</td>
<td>33 to 34</td>
<td>2</td>
</tr>
<tr>
<td>Herculano-Houzel et al.</td>
<td>2008</td>
<td>35 to 40</td>
<td>6</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2008</td>
<td>41 to 44</td>
<td>4</td>
</tr>
<tr>
<td>Kabat et al.</td>
<td>2008</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>Hedenström</td>
<td>2008</td>
<td>46 to 49</td>
<td>4</td>
</tr>
<tr>
<td>McCoy &amp; Gillooly</td>
<td>2008</td>
<td>50 to 53</td>
<td>4</td>
</tr>
<tr>
<td>Boback &amp; Guyer</td>
<td>2008</td>
<td>54</td>
<td>1</td>
</tr>
<tr>
<td>Yopak &amp; Montgomery</td>
<td>2008</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>Plavcan &amp; Ruff</td>
<td>2008</td>
<td>56 to 171</td>
<td>116</td>
</tr>
<tr>
<td>Ashwell</td>
<td>2008</td>
<td>172 to 175</td>
<td>4</td>
</tr>
<tr>
<td>Hofmann et al.</td>
<td>2008</td>
<td>176 to 179</td>
<td>4</td>
</tr>
<tr>
<td>Hall</td>
<td>2008</td>
<td>180 to 181</td>
<td>2</td>
</tr>
<tr>
<td>Cuc &amp; Ninomiya</td>
<td>2007</td>
<td>182 to 184</td>
<td>3</td>
</tr>
<tr>
<td>Hughes &amp; Hughes</td>
<td>2007</td>
<td>185 to 186</td>
<td>2</td>
</tr>
<tr>
<td>Yopak et al.</td>
<td>2007</td>
<td>187</td>
<td>1</td>
</tr>
<tr>
<td>Macrini et al.</td>
<td>2007</td>
<td>188 to 193</td>
<td>6</td>
</tr>
<tr>
<td>Rubensson et al.</td>
<td>2007</td>
<td>194 to 196</td>
<td>3</td>
</tr>
<tr>
<td>Marbà et al.</td>
<td>2007</td>
<td>197 to 199</td>
<td>3</td>
</tr>
<tr>
<td>Chalfin et al.</td>
<td>2007</td>
<td>200 to 202</td>
<td>3</td>
</tr>
<tr>
<td>Alerstam et al.</td>
<td>2007</td>
<td>203 to 210</td>
<td>8</td>
</tr>
<tr>
<td>Witzmann &amp; Scholz</td>
<td>2007</td>
<td>211 to 232</td>
<td>22</td>
</tr>
<tr>
<td>Lemelin &amp; Jungers</td>
<td>2007</td>
<td>233</td>
<td>1</td>
</tr>
<tr>
<td>Pillay &amp; Manger</td>
<td>2007</td>
<td>234-241</td>
<td>8</td>
</tr>
<tr>
<td>Price &amp; Enquist</td>
<td>2007</td>
<td>242-352</td>
<td>111</td>
</tr>
<tr>
<td>Savage et al.</td>
<td>2007</td>
<td>353-361</td>
<td>9</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Year</td>
<td>Pages</td>
<td>Count</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>Torres &amp; Vanni</td>
<td>2007</td>
<td>362-367</td>
<td>6</td>
</tr>
<tr>
<td>Herculano-Houzel et al.</td>
<td>2007</td>
<td>368-380</td>
<td>13</td>
</tr>
<tr>
<td>Helgen et al.</td>
<td>2006</td>
<td>381</td>
<td>1</td>
</tr>
<tr>
<td>Marroig</td>
<td>2006</td>
<td>382-385</td>
<td>4</td>
</tr>
<tr>
<td>Iwaniuk et al.</td>
<td>2006</td>
<td>386-389</td>
<td>4</td>
</tr>
<tr>
<td>Mortola &amp; Limoges</td>
<td>2006</td>
<td>390</td>
<td>1</td>
</tr>
<tr>
<td>Tsahar et al.</td>
<td>2006</td>
<td>391-394</td>
<td>4</td>
</tr>
<tr>
<td>Nummela &amp; Sánchez-Villagra</td>
<td>2006</td>
<td>395-403</td>
<td>9</td>
</tr>
<tr>
<td>Snively et al.</td>
<td>2006</td>
<td>404-405</td>
<td>2</td>
</tr>
<tr>
<td>Sherwood et al.</td>
<td>2006</td>
<td>406-407</td>
<td>2</td>
</tr>
<tr>
<td>Serrano-Meneses &amp; Székely</td>
<td>2006</td>
<td>408</td>
<td>1</td>
</tr>
<tr>
<td>Manger</td>
<td>2006</td>
<td>409-418</td>
<td>10</td>
</tr>
<tr>
<td>Withers et al.</td>
<td>2006</td>
<td>419-421</td>
<td>3</td>
</tr>
<tr>
<td>Raihani et al.</td>
<td>2006</td>
<td>422</td>
<td>1</td>
</tr>
<tr>
<td>Miller &amp; Eadie</td>
<td>2006</td>
<td>423</td>
<td>1</td>
</tr>
<tr>
<td>Sun et al.</td>
<td>2006</td>
<td>424-436</td>
<td>13</td>
</tr>
<tr>
<td>Makarieva et al.</td>
<td>2005</td>
<td>437</td>
<td>1</td>
</tr>
<tr>
<td>Schmidt</td>
<td>2005</td>
<td>438-442</td>
<td>5</td>
</tr>
<tr>
<td>Niven &amp; Scharlemann</td>
<td>2005</td>
<td>443-446</td>
<td>4</td>
</tr>
<tr>
<td>Thomsen &amp; Wernberg</td>
<td>2005</td>
<td>447</td>
<td>1</td>
</tr>
<tr>
<td>White</td>
<td>2005</td>
<td>448-449</td>
<td>2</td>
</tr>
<tr>
<td>Riskin et al.</td>
<td>2005</td>
<td>450-453</td>
<td>4</td>
</tr>
<tr>
<td>Carbone et al.</td>
<td>2005</td>
<td>454</td>
<td>1</td>
</tr>
<tr>
<td>Lane et al.</td>
<td>2004</td>
<td>455</td>
<td>1</td>
</tr>
<tr>
<td>Van der Meji &amp; Bout</td>
<td>2004</td>
<td>456-461</td>
<td>6</td>
</tr>
<tr>
<td>Makarieva et al.</td>
<td>2004</td>
<td>462</td>
<td>1</td>
</tr>
<tr>
<td>Radloff &amp; Du Toit</td>
<td>2004</td>
<td>463-464</td>
<td>2</td>
</tr>
<tr>
<td>McKechnie &amp; Wolf</td>
<td>2004</td>
<td>465-466</td>
<td>2</td>
</tr>
<tr>
<td>Klaassen &amp; Nolet</td>
<td>2008</td>
<td>467-469</td>
<td>3</td>
</tr>
<tr>
<td>Anderson &amp; Jetz</td>
<td>2005</td>
<td>470-471</td>
<td>2</td>
</tr>
</tbody>
</table>
Appendix C  Detailed Simulation Procedures

Technical details omitted for Monte Carlo simulation with empirical parameters and weighted average model are provided below.

Simulation Procedure with Empirical Parameters

This section corresponds to Error Distribution Determines Best Method for Fitting Power-Laws in the main text. Each empirical dataset was assumed to have a multiplicative log-normal error structure, and $a$, $b$, and $\sigma$ were estimated with LR (with $\sigma$ estimated as the standard deviation of the residuals). $10^5$ independent simulations were carried out using the estimated parameters. Each simulation was analyzed with both LR and NLR. Next, an additive normal error structure was assumed for the same dataset, the parameters were estimated with NLR, and the Monte Carlo simulation and analysis procedures were repeated. While the assumption of log-normal error always generated feasible datasets with positive $y$ values, assuming normal error can lead to unrealistic negative $y$ values under some parameterizations. This failure of certain empirical parameterizations to produce biologically acceptable simulated data indicated that the assumed normal error structure was inappropriate for the original dataset. Thus we did not analyze 232 empirical datasets for which more than 10% of simulated datasets included negative $y$ values when the error structure was assumed to be normal; for the remaining empirical datasets, we discarded any simulated dataset with negative $y$ values.

Simulation Procedure for Model Averaging

This section corresponds to Model Averaging: An Alternative Approach When Error Form Is Uncertain in the main text. To test model averaging as an approach for
dealing with uncertain error structure we repeated the empirically motivated simulation process (described above in Simulation Procedure with Empirical Parameters) assuming first log-normal error and then normal error with $10^5$ simulations for each empirical dataset. Each simulated dataset was analyzed with both LR and NLR. Weighted point estimates for $a$ and $b$ were obtained as the weighted average of the point estimates from the two models using AICc weights computed from AICc values.

Because no simple analytical form exists for CIs in the weighted model, CIs for $a$ and $b$ were computed with bootstrapping. Due to computational limitation, the bootstrapping process was carried out with 400 additional independent simulations for each empirical dataset. For each simulated dataset 500 bootstrap samples were constructed, and AICc-weighted estimates for $a$ and $b$ were calculated for each sample. A 95% CI for each parameter was then determined as the range between the 2.5% and 97.5% quantiles of all bootstrap estimates.
library(nlrwr)
library(boot)

## Function to carry out the analytical process described in General Guidelines
## Input: x - vector, explanatory variable
## y - vector, response variable
## CI_boot - whether confidence intervals should be computed using bootstrapping
## for model averaging. Default is TRUE.
## output_plot - whether a scatter plot with fitted relationship is desired.
## default is FALSE.

## Output: method - method used in analysis
## a, b - estimated parameters
## a_confint, b_confint - 95% confidence intervals for a & b
## (optional if method is "Model Averaging")

power_analysis = function (x, y, CI_boot = TRUE, output_plot = FALSE){

  ## Step 1: Likelihood analysis
  model_lr = lm(log(y) ~ log(x))
  a_lr = exp(coef(summary(model_lr))[1, 1])
  b_lr = coef(summary(model_lr))[2, 1]
  sd_lr = sd(log(y) - (log(a_lr) + b_lr * log(x)))

  model_nlr = nls(y ~ a1 * x ^ a2, start = list(a1 = a_lr, a2 = b_lr),
                  control = nls.control(maxiter = 2000, warnOnly = TRUE))
  a_nlr = coef(summary(model_nlr))[1, 1]
  b_nlr = coef(summary(model_nlr))[2, 1]
sd_nlr = sd(y - a_nlr * x ^ b_nlr)

l_logn = sum(log(dlnorm(y, log(a_lr * x ^ b_lr), sd_lr)))

l_norm = sum(log(dnorm(y, a_nlr * x ^ b_nlr, sd_nlr)))

n = length(x)

k = 3

AICc_logn = 2 * k - 2 * l_logn + 2 * k * (k + 1) / (n - k - 1)

AICc_norm = 2 * k - 2 * l_norm + 2 * k * (k + 1) / (n - k - 1)

delta_AICc = AICc_norm - AICc_logn

writeLines(paste("AICc_logn: ", AICc_logn, "nAICc_norm: ", AICc_norm))

w_logn = exp(-(AICc_logn - min(AICc_logn, AICc_norm)) / 2)

w_norm = exp(-(AICc_norm - min(AICc_logn, AICc_norm)) / 2)

weight_logn = w_logn / (w_logn + w_norm)

weight_norm = w_norm / (w_logn + w_norm)

## Step 2a: Analysis with NLR

if (delta_AICc < -2) {

    writeLines("The assumption of additive normal error is better supported. Proceed with NLR.")

    method = "NLR"

    a = a_nlr

    b = b_nlr

    a_confint = confint2(model_nlr)[1,]

    b_confint = confint2(model_nlr)[2,]

}
### Step 2b: Analysis with LR

```r
else if (delta_AICc > 2) {
    writeLines("The assumption of multiplicative log-normal error is better supported. Proceed with LR.")

    method = "LR"
    a = a_lr
    b = b_lr
    a_confint = confint(model_lr)[1,]
    b_confint = confint(model_lr)[2,]
}
```

### Step 2c: Analysis with model averaging

```r
else {
    writeLines("The two error distributions have similar support. Proceed with model averaging.")

    method = "Model Averaging"
    a = a_lr * weight_logn + a_nlr * weight_norm
    b = b_lr * weight_logn + b_nlr * weight_norm
    if (!CI_boot) {
        a_confint = NA
        b_confint = NA
    }
    else {
        boot.est = function(dat, indices) {
            dat.sub = dat[indices,]
            names(dat.sub) = c("x", "y")
            model.lr = lm(log(y) ~ log(x), dat = dat.sub)
```
a.lr = \exp(\text{coef}(\text{summary(model.lr)})[1, 1])

b.lr = \text{coef}(\text{summary(model.lr)})[2, 1]

sd.lr = \text{sd}(\log(\text{dat.sub}$y) - (\log(a.lr) + b.lr \times \log(\text{dat.sub}$x)))

a.lr.CI = \text{confint(model.lr)}[1,]

b.lr.CI = \text{confint(model.lr)}[2,]

model.nlr = \text{nls}(y \sim a1 \times x \wedge a2, \text{start} = \text{list}(a1 = a.lr, a2 = b.lr), \text{dat} = \text{dat.sub},

\text{control} = \text{nls.control(maxiter = 2000, warnOnly = TRUE)})

a.nlr = \text{coef}(\text{summary(model.nlr)})[1, 1]

b.nlr = \text{coef}(\text{summary(model.nlr)})[2, 1]

sd.nlr = \text{sd}(\text{dat.sub}$y - a.nlr \times \text{dat.sub}$x \wedge b.nlr)

a.nlr.CI = \text{confint2(model.nlr)}[1,]

b.nlr.CI = \text{confint2(model.nlr)}[2,]

l.logn = \text{sum}(\log(\text{dlnorm(dat.sub}$y, \log(a lr \times \text{dat.sub}$x \wedge b lr), \text{sd lr)))

l.norm = \text{sum}(\log(\text{dnorm(dat.sub}$y, a nlr \times \text{dat.sub}$x \wedge b nlr, \text{sd nlr}))

AICc.logn = 2 \times k - 2 \times l.logn + 2 \times k \times (k + 1) / (n - k - 1)

AICc.norm = 2 \times k - 2 \times l.norm + 2 \times k \times (k + 1) / (n - k - 1)

AICc.min = \min(AICc.logn, AICc.norm)

weight.logn = \exp(-(AICc.logn - AICc.min)/2)

weight.norm = \exp(-(AICc.norm - AICc.min)/2)

logn.w = weight.logn / (weight.logn + weight.norm)

norm.w = weight.norm / (weight.logn + weight.norm)

a.boot = a.lr \times \text{logn.w} + a.nlr \times \text{norm.w}

b.boot = b.lr \times \text{logn.w} + b.nlr \times \text{norm.w}
return(c(a.boot, b.boot))

}
dat.boot = boot(data = as.data.frame(cbind(x, y)), statistic = boot.est, R = 1000)
a_confint = boot.ci(dat.boot, index = 1, type = "perc")$perc[4:5]
b_confint = boot.ci(dat.boot, index = 2, type = "perc")$perc[4:5]
}
}
writeLines(paste("a: ", a, " nb: ", b))

if (output_plot) {
par(mfrow = c(1, 2), oma = c(0, 4, 2, 0))
plot(x, y, log = "xy", xlab = "x", ylab = "y", pch = 20, main = "Logarithmic Scale")
curve(a * x^b, add = TRUE, lty = "dashed")
plot(x, y, xlab = "x", ylab = "y", pch = 20, main = "Arithmetic Scale")
curve(a * x^b, add = TRUE, lty = "dashed")
title(main = paste("Fitting Power-Law Data with ", method), outer = TRUE)
}

return (list(method = method, a = a, b = b, a_confint = a_confint, b_confint = b_confint))
}